CESifo WORKING PAPERS

10233
2023

January 2023

Strategic Behavior with Tight, Loose and Polarized Norms

Eugen Dimant, Michele Gelfand, Anna Hochleitner, Silvia Sonderegger

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 10233*

Strategic Behavior with Tight, Loose and Polarized Norms

Abstract

Descriptive norms – the behavior of other individuals in one's reference group – play a key role in shaping individual decisions. When characterizing the behavior of others, a standard approach in the literature is to focus on average behavior. In this paper, we argue both theoretically and empirically that not only averages but also the shape of the whole distribution of behavior can play a crucial role in how people react to descriptive norms. Using a representative sample of the U.S. population, we experimentally investigate how individuals react to strategic environments that are characterized by different distributions of behavior, focusing on the distinction between tight (i.e., characterized by low behavioral variance), loose (i.e., characterized by high behavioral variance), and polarized (i.e., characterized by u-shaped behavior) environments. We find that individuals indeed strongly respond to differences in the variance and shape of the descriptive norm they are facing: loose norms generate greater behavioral variance and polarization generates polarized responses. In polarized environments, most individuals prefer extreme actions – which expose them to considerable strategic risk – to intermediate actions that minimize such risk. Importantly, we also find that relative to tight environments, in polarized and loose environments, personal traits and values play a larger role in determining actual behavior. This provides important insights into how individuals navigate environments that contain strategic uncertainty.

JEL-Codes: C910, D010.

Keywords: cooperation, descriptive norms, variance, peer effects.

Eugen Dimant
University of Pennsylvania
Philadelphia / PA / USA
edimant@sas.upenn.edu

Michele Gelfand
Stanford University
Stanford / CA / USA
gelfand1@stanford.edu

Anna Hochleitner
University of Nottingham / United Kingdom
anna.hochleitner@nottingham.ac.uk

Silvia Sonderegger
University of Nottingham / United Kingdom
silvia.sonderegger@nottingham.ac.uk

January 16, 2023
We are grateful to participants at the 2022 ASSA in Boston, the FAIR workshop organized by the Norwegian School of Economics, the 2022 ICSD in Copenhagen, the 2022 SEA meeting in Basel as well as seminars in Essex, Konstanz, Montpellier, Hamburg, Leicester, Nottingham, Oxford, and the WZB Berlin. We also thank Zvonimir Baši¢, Valerio Capraro, Laura Gee, Steffen Huck, Felix Kölle, Jonathan Schulz, Alexander Vostroknutov, and Max Winkler for helpful comments on earlier drafts of this paper. We thank our research assistant Nikolas Alves Da Costa Silva for his help. Financial support from Stanford University and the German Research Foundation (DFG) under Germany's Excellence Strategy – EXC 2126/1– 390838866 is gratefully acknowledged. This project was pre-registered at https://aspredicted.org/pm7fu.

1 Introduction

Descriptive norms - the behavior of other individuals in one's reference group - play a key role in shaping individual decisions.[1] When characterizing the behavior of others, a common approach in the existing social science literature is to focus on *mean* or *modal* behavior.[2] In this paper, we argue that focusing primarily on mean or modal behavior fails to account for important features – the *variance* or *shape* of the distribution of behavior – that may play a vital role in how people react to a descriptive norm in the presence of strategic interactions.

Consider a collective action problem and suppose that individuals in a given society contribute an average of 2 out of a maximum of 4 tokens. This may reflect a situation where either *everyone* contributes 2, where each contribution level is selected by an *equal share* of the population, or where *half* of the population contributes nothing and half contributes everything. Now consider an agent who interacts with a random partner in one of these societies. The agent does not know their partner's contribution but knows that it is drawn from the distribution that characterizes the descriptive norm in that society. How much should the agent contribute? A mean-focused approach suggests that their contribution should be the same, independently of which descriptive norm they are confronted with. Crucially, however, although these scenarios generate the same average contribution, they clearly depict very different social environments.

The role of variance in characterizing descriptive norms is emphasized by Gelfand et al. (2011) and Gelfand (2021), who distinguish between *tight* and *loose* norms, arguing that this distinction can help to understand systematic differences across cultures (see also Winkler, 2021). Tight cultures are characterized by well-defined behavior, while loose cultures show a pattern of greater behavioral variance. Implicit in this approach is the idea that, when faced with a loose norm, people's reactions exhibit more variation and vice versa for tight norms, generating *multiple equilibria* that can be expressed in different cultural characteristics.

The focus of this paper is to shed light on how people react to different features of descriptive norms in a context that is arguably one of the cornerstones of human cooperation and which is ubiquitous in all environments involving social interactions: public goods provision.[3] The existing literature has documented substantial heterogeneity in the distribution of contributions in public good dilemmas across cultures (Henrich et al., 2001a,b; Gächter et al., 2010), making it a particularly interesting case study for examining the effects of different descriptive norms. In addition to variance, we investigate another important feature that characterizes the distribution of behavior, namely its shape (unimodal versus u-shaped or polarized). Understanding how

[1] E.g., Cialdini and Trost (1998); Bicchieri (2005); Fisman and Miguel (2007); Perez-Truglia and Cruces (2017); Fehr and Schurtenberger (2018); Bicchieri et al. (2022b).

[2] See e.g. Kandel and Lazear (1992); Grout et al. (2015); Feldhaus et al. (2019). This of course does not imply that other aspects of distribution have been entirely neglected (see e.g., Bicchieri and Xiao (2009); Krupka and Weber (2013); Adriani and Sonderegger (2019); Michaeli and Spiro (2017, 2015). However, as we will argue, the literature currently lacks a systematic investigation of how the variance and shape of the descriptive norm affect individual behavior in strategic environments.

[3] It is important to stress that here we use the term "descriptive norm" to indicate the distribution of behavior, differently from the use of the term which can be found elsewhere (Cialdini et al., 1990; Bicchieri, 2005) where it indicates "what people commonly do." By focusing on descriptive norms, we abstract from injunctive norms that look at "what people commonly approve or disapprove of" (Bicchieri, 2005). Our approach allows us to isolate the effect of one precise aspect of norms.

people react to polarized environments has become increasingly important, especially in the face of increasing political polarization and its detrimental societal outcomes.[4]

We consider a one-shot public good game where two individuals need to sacrifice some of their self-interest to further joint welfare. It is well known that, when deciding on their contributions in social dilemmas, people exhibit strong social preferences in the form of conditional reciprocity (see e.g. Gächter et al., 2010, 2017).[5] For example, when facing a high contributor, they tend to react by choosing a high contribution themselves, while when facing a low contributor they react by contributing little. However, people typically do not know their co-player's contribution when making their choice. Instead, they face *strategic uncertainty*. All they know is that their co-player's contribution will be drawn from a probability distribution that corresponds to the prevailing descriptive norm. It is therefore important to understand how the features of this distribution (its mean, but also its variance and shape) affect the individual's optimal response.

Why should individuals react to the variance and shape of the descriptive norm they are facing? Our premise is that different distributions of cooperative behavior generate different degrees of strategic uncertainty. In very tight environments, where there is low variance in behavior, strategic uncertainty is minimal, while in loose or polarized environments where there is high variance, the behavior of other individuals is less predictable and strategic uncertainty is substantial. In this paper, we investigate both theoretically and empirically how people respond to different levels of strategic uncertainty when they are strongly motivated by reciprocal concerns. People may focus only on the mean of the distribution they face – as suggested by the *mean-based* approach – or they may react to both the mean and the variance/shape of the distribution – as implied by the *multiple equilibria* approach described above. As we explain below, our findings strongly support the latter view.

Using a representative sample of the U.S. population, we examine our research question experimentally through the lens of a well-powered (N=1203) and pre-registered study.[6] We do so by introducing a variant of the established public goods game (PGG) with two players as used by Gächter et al. (2017). Players receive a number of tokens at the beginning of the game and can decide to keep them for themselves or invest them in a public good that is then multiplied by a positive factor and shared equally among both players. The experiment is divided into two parts. In Part I, we use the ABC strategy method (Fischbacher et al., 2001; Gächter et al., 2017) to elicit the participants' underlying cooperative propensity, as well as their beliefs about their co-player's contribution. In Part II we then present the participants with the distribution from which the co-player's contribution will be drawn. In a between-subject design, we implement six different treatments that vary the mean (high/low) and variance/shape (tight/loose/polarized) of the co-player's distribution.

[4]See Iyengar and Westwood (2015); McConnell et al. (2018); Iyengar et al. (2019); Enke (2020); Bauer et al. (2022); Dimant et al. (2022); Dimant (2022b); Ren et al. (2023); Robbett and Matthews (2022); Nunn (2022); Callander and Carbajal (2022). What we call polarized norms can be equivalently interpreted as reflecting the coexistence of two different norms in the population. Thus, while we use the term "polarized descriptive norm" to indicate a polarized distribution of behavior, it should be clear that this is simply a semantic choice and should not be seen as conflicting with a "dual norm" interpretation.

[5]See also (Kölle and Quercia, 2021) who show that matching others' contributions is perceived by subjects as being the most appropriate thing to do.

[6]See https://aspredicted.org/pm7fu.

Our results confirm that considering the mean as a sufficient statistic when describing norms may provide an incomplete picture. We find that exposure to norms that share the same average behavior generates very different responses depending on their precise nature (tight, loose, polarized). In line with our theoretical examination, the distribution of participants' contributions roughly replicates the descriptive norm they are confronted with. When the descriptive norm they face is tight, the participants' contributions are narrowly distributed. When the descriptive norm they face is loose, the participants' contributions are spread out. Similarly, when confronted with a polarized descriptive norm, the participants' reactions are concentrated at the extremes of the distribution: they either choose to contribute a lot or very little. Interestingly, we find that in polarized environments a large share of participants choose the maximum contribution, thus exposing themselves to considerable strategic risk: depending on who they are matched with, their contribution will either perfectly match that of their co-player or will be as far as possible from it. These participants could shield themselves from strategic risk by choosing a middle-of-the-road contribution, but they do not. Rather than focusing on minimizing strategic risk, they appear to be primarily concerned with not letting high contributors down. This suggests that, in the social domain, people are tolerant of and may even embrace risk.

Our results also point to an interaction between strategic uncertainty and the role of personal traits: When strategic uncertainty is high, people turn to their preferences and personal values to decide how to behave (see also Elster and Gelfand, 2021). We find that *sucker aversion* (the aversion to contributing more than the co-player), *free-riding aversion* (the aversion to contributing less than the co-player), and *personal values* (what people perceive to be the "right thing to do") play a larger role in determining behavior when individuals are confronted with a loose/polarized environment compared to a tight one. This underlines the importance of considering personal values in addition to descriptive norms when making behavioral predictions (see e.g. Bicchieri, 2005; Bicchieri and Dimant, 2019; Capraro et al., 2019; Bašić and Verrina, 2020) and has clear practical implications, e.g. for policy-makers interested in assessing the role of norms and values in influencing behavior (see e.g. Barr et al., 2020, for a discussion on the importance of distinguishing norms from personal values for facilitating behavioral change).

Taken together, our findings contribute to the existing literature in various ways. Firstly, we add to a growing body of research that explores the effect of descriptive norms on individual behavior. Many studies have shown in different contexts that providing information on what other people did in a given situation influences individual decisions. This has been found in both nonstrategic settings such as dictator games (Bicchieri and Xiao, 2009), voluntary payments (Shang and Croson, 2009; Feldhaus et al., 2019) and donations to charities (Dimant, 2019; Bicchieri et al., 2022a), as well as in strategic interactions such as public goods provision (Chaudhuri et al., 2006; Kerr et al., 2009; Chen et al., 2010). These studies typically communicate information about others using mean or modal behavior. Consistent with this literature, our study also finds that differences in means have a significant effect on subsequent decisions. However, we extend these findings by exploring differences between the whole distribution.

Our investigation is also related to the work by d'Adda et al. (2020), in which dictators are shown different distributions of normative views taken from a previous study (baseline, low mean, and high variance) before selecting their action. In line with our results, they provide evidence indicating that in the high variance treatment the variance of the dictator contributions is higher.[7] Importantly, however, the game they consider does not feature any strategic interaction and the distributions they observe concern the *normative* views of others. In contrast, the setup we analyze is one within which individuals are confronted with strategic uncertainty. Participants are informed about the distribution from which the contribution of their co-player is drawn. Our focus is therefore on the reaction of participants to the *behavior* of others – the descriptive norm – in a setting where participants are known to exhibit reciprocal motives. The psychological mechanism behind our results is therefore fundamentally different from d'Adda et al. (2020). Moreover, we consider a wider range of distributions and systematically vary not only their variance but also their shape (unimodal or polarized). This allows for a richer investigation of behavioral responses, which we will isolate and analyze in turn.

We also add to the literature that stresses conditional cooperation as a powerful motive for explaining contributions to a public good (see e.g. Fischbacher and Gächter, 2010; Bowles and Gintis, 2013; Gächter et al., 2017). A number of studies examine how heterogeneity in contributions in one's interacting group affects (conditional) cooperation - Chaudhuri et al. (2006); Kerr et al. (2009); Croson (2007); Wolff (2017). Using the strategy method, Cheung (2014) and Hartig et al. (2015) consider a setup where individuals interact in groups and the actions of all group members are known, showing that people react not only to averages but the whole profile of individual contribution. Our results extend these findings by studying the effects of behavioral heterogeneity in a setup characterized by *strategic uncertainty*, where participants do not know their partner's actions in advance.

A final contribution of our work is that we develop and test a novel norm elicitation approach that allows us to measure not only beliefs about the *mean* as is the case in established elicitation methods such as Bicchieri and Chavez (2010) or Krupka and Weber (2013), but also about the entire distribution in an incentive-compatible way (for a discussion, see Dimant, 2022a). We provide a fine-grained measurement tool to develop a better understanding of descriptive norms and their impact on behavior that can be used in future research.[8]

The remainder of this paper is structured as follows. In Section 2, we outline our theoretical framework and hypotheses. Section 3 describes the experimental design. Section 4 presents the empirical results, while Section 5 provides a discussion and concludes.

2 Theoretical framework

It is well known that, in strategic environments, reciprocity plays an important role in determining an individual's choice of action (Fehr and Gächter, 2000). The literature on public goods games extensively documents the presence of reciprocity motives (see e.g. Fischbacher and Gächter, 2010; Bowles and Gintis, 2013; Gächter et al., 2017). When faced with a high contributor, participants contribute a lot, while when faced with a low contributor, they contribute little. This shows that individuals are strongly concerned with matching the behavior of others, and

[7]This result however disappears in the treatment where the beliefs of participants are also elicited.
[8]The software eliciting norm-related beliefs using the distribution builder can be downloaded here.

substantiates our investigation on the effect of descriptive norms.

The underlying idea is that, consistent with their concerns for reciprocity, individuals incur a psychological loss whenever their contribution differs from that of their co-player. Consequently, they adapt their behavior to the behavior of their co-player. At the same time, when engaging in everyday interactions people cannot fully anticipate what their counterparts will do, in spite of being aware of the typical distribution of behavior within society (the descriptive norm). Since their co-player's contribution is unobserved at the time when they choose their action, agents are exposed to *strategic uncertainty*. Their choice needs to trade off the risk of contributing too little (relative to their co-player) and the risk of contributing too much. These competing factors determine the optimal contribution for an individual when confronted with a distribution of co-player's actions.

As remarked in the introduction, a standard approach in economics is to use the mean of the distribution of others' behavior as a sufficient statistic to determine an individual's optimal reaction. This would suggest that people are indifferent to the variance/shape of the distribution. An alternative view is that people *do* react to the variance/shape of the distribution they are facing and that their reactions mimic the initial distribution, generating multiple equilibria.

In what follows, we show that these competing views can be captured by employing two canonical loss functions widely used in statistics (see e.g. DeGroot, 2005) as well as economics to model the psychological cost of a mismatch between own and co-player's contribution: (i) *quadratic*: cost is proportional to the square of the difference between contributions, and (ii) *absolute value*: cost is proportional to the absolute value of the difference between contributions.

Quadratic loss functions are commonly used, for instance, in models of conformity or coordination (see e.g. Kandel and Lazear, 1992; Grout et al., 2015). The absolute value loss function is widely used following a seminal contribution by Fehr and Schmidt (1999).

Let x_i denote one's own contribution, x_j one's co-player's contribution and X the endowment. We are interested in a setup where individuals do not observe their co-player's behavior before selecting their action, but know that the action of their co-player is drawn from a distribution $f(x)$ on $[0, \overline{x}]$ with mean μ and variance σ^2. Note that, for ease of exposition, in this analysis we focus on continuous approximations of the discrete distributions we use in our experiment, which are depicted in Figure 1 of Section 3. All proofs can be found in Appendix A.

2.1 Case (i): Individuals react only to the mean

Consider the following stylized model of reciprocal preferences:

$$u_i = X - x_i + \gamma(x_i + x_j) - \frac{\eta_i}{2}(x_i - x_j)^2 - \frac{\delta_i}{2}(x_i - x_i^a)^2. \tag{1}$$

where $X - x_i + \gamma(x_i + x_j)$ is material payoff (for some $\frac{1}{2} < \gamma < 1$), $\eta_i \geq 0$ parametrizes i's reciprocity concerns, $x_i^a \in [0, \overline{x}]$ is what i considers the "right thing to do" (their *personal value*) and $\delta_i \geq 0$ captures the importance that i ascribes to acting in accordance to their personal value.[9] The first quadratic term in (1) captures the desire to minimize the psychological loss incurred whenever the player's contribution differs from that of the co-player ("mismatch loss"), while the last quadratic term in (1) models the psychological cost incurred by i when deviating from their personal value. Each individual i selects x_i to maximize their expected utility, where the expectation is taken with respect to x_j. We denote i's optimal contribution as x_i^*.

Proposition 1: *When utility is given by (1), we have (i) $x_i^* = 0$ if $\eta_i < (1-\gamma-\delta_i x_i^a)/\mu$, (ii) $x_i^* = [\eta_i\mu + \delta_i x_i^a - (1-\gamma)]/(\eta_i+\delta_i)$ otherwise.*

Intuitively, when the mismatch loss from selecting a contribution that is different from the co-player's is quadratic, individuals are *averse to strategic risk.* They choose their contribution to minimize the strategic risk they are exposed to. The optimal solution to this problem indexes i's contribution to μ, the co-player's mean contribution. This ensures that the difference between x_i and x_j is never too large. Crucially, it implies that i's choice *only* depends on $f(x)$ through μ, and is independent of the other features of the distribution of co-player's behavior.

2.2 Case (ii): Individuals react to the whole distribution

Suppose now that utility is

$$u_i = X - x_i + \gamma(x_i + x_j) - \alpha_i (x_i - x_j)\,|_{x_j<x_i} - \beta_i (x_j - x_i)\,|_{x_j>x_i} - \frac{\delta_i}{2}(x_i - x_i^a)^2. \qquad (2)$$

This utility function differs from (1) in that the mismatch loss incurred by individuals is proportional to the *absolute value* of the difference between their contribution and that of their co-player. The parameter $\alpha_i \geq 0$ (resp., $\beta_i \geq 0$) measures the marginal disutility obtained from selecting a contribution that exceeds (resp., is lower than) the co-player's contribution.

Proposition 2: *Let $\phi_i \equiv \beta_i - (1-\gamma) + \delta_i x_i^a$. When utility is given by (2), we have (i) $x_i^* = 0$ if $\phi_i \leq 0$, (ii) x_i^* satisfies $\delta_i x_i^* + F(x_i^*)(\alpha_i+\beta_i) = \phi_i$ otherwise.*

To fix ideas, consider the simple case where $\delta_i = 0$ so that, when interior, x_i^* satisfies $F(x_i^*) = \varphi_i$ defined as $\varphi_i \equiv \frac{\phi_i}{\alpha_i+\beta_i}$. Figure 1 represents the function $F(x)$ for the case of (i) single-peaked distributions and (ii) polarized (u-shaped) distributions. In panel (i), the solid line represents a distribution with a smaller variance compared to the dashed line. The horizontal straight lines represent φ_i.

As can be seen from Figure 1, the point where $F(x)$ and φ_i cross depends on the nature of the distribution of co-player contributions. For instance, when $f(.)$ is polarized, $F(x)$ is steep at the extremes and flat in the middle. This implies that, typically, $F(x)$ and φ_i will cross when x takes extreme values – either very low or very high (panel (ii) of the Figure 1 illustrates the latter possibility). When facing a polarized distribution, individuals thus exhibit strategic risk-taking behavior: they prefer to take a gamble and risk ending up in a completely mismatched position

[9] In the main body we focus on the simple case where the psychological loss from deviating from one's personal norm is quadratic. In Appendix A we discuss the case where this loss depends on the absolute value. Finally, it is worth noting that, although the disutility from contributing a different amount from the co-player will typically depend on the degree of intentionality in the co-player's action, this is immaterial here since in our design intentionality is the same across all treatments.

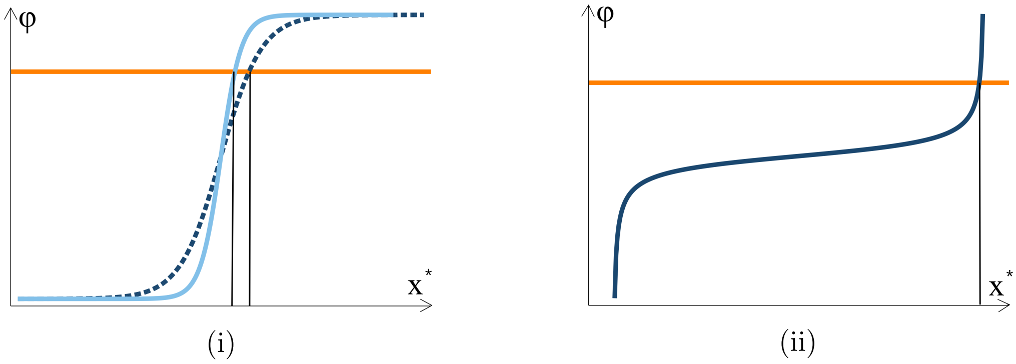


Figure 1: F(x) for single peaked (i) and polarized (ii) distributions

vis-à-vis their co-player rather than opting for a "middle of the road" contribution level which would minimize risk.

In contrast, when the distribution of co-player contributions is single-peaked, $F(x)$ is flat at the extremes and steep in the middle. Consequently, $F(x)$ and φ_i will tend to cross when x takes intermediate values. As the variance of $f(.)$ increases, though, x_i^* will tend to become progressively more extreme, as can be seen by comparing the solid and the dashed lines in the panel (i) of Figure 1.

The following result formalizes the notion that, as the variance of co-player contribution increases, individuals tend to select more extreme contributions. Consider two distributions f_0 and f_1 with the same mean μ and suppose that f_0 is single-crossing stochastic dominant over f_1 (Machina and Pratt, 1997) so that, for some $\widehat{x} \in (0, \overline{x})$, the following holds: $F_1(x) > F_0(x)$ for $x < \widehat{x}$ and $F_1(x) < F_0(x)$ for $x > \widehat{x}$. In our experiment, this is satisfied for all pairwise comparisons of descriptive norms sharing the same mean (see Appendix B, Figure B.1), with $\widehat{x} = 2$ in all cases. Note that this condition implies that f_1 is a mean-preserving spread of f_0. Denoting the optimal contribution under f_k as x_{ik}^*, the following holds.

Corollary 1: *(i) For all individuals i for whom* $x_{i0}^* \in (0, \widehat{x}) : x_{i0}^* > x_{i1}^*$. *(ii) For all individuals i for whom* $x_{i0}^* \in (\widehat{x}, \overline{x}) : x_{i0}^* < x_{i1}^*$.

In other words, when individuals are confronted with a distribution that is more spread out, their responses are also more spread out. People who choose a low contribution when facing f_0 choose an even *lower* contribution when confronted with the more spread out distribution f_1. Vice versa, those who choose a high contribution when facing f_0 choose an even *higher* contribution when confronted with f_1. We now look at the role of individual traits, α_i, β_i and x_i^a, in determining individual contributions.

Corollary 2 *When interior, optimal contributions are decreasing in* α_i, *increasing in* β_i *and increasing in* x_i^a.

Next, we compare the effect of a change in individual traits on the optimal contribution when the individual is confronted with f_0 vs the more spread out distribution f_1. The underlying question is whether the nature of the descriptive norm would affect the extent to which personal traits influence contributions.

Corollary 3 *Consider* $x'' > \widehat{x} > x'$. *The parameter shift (either in* α_i, β_i *or* x_i^a*) needed to* *generate a change in optimal contribution from x' to x'' is larger under f_0 than under f_1.*

Figure 2 illustrates the result for a tight versus a polarized distribution, in the easy-to-depict case where $\delta_i = 0$. The change in φ needed to generate a shift in optimal contribution from x' to x'' is much larger when the individual faces a tight distribution (panel (i)) compared to the case of a polarized distribution (panel (ii)).

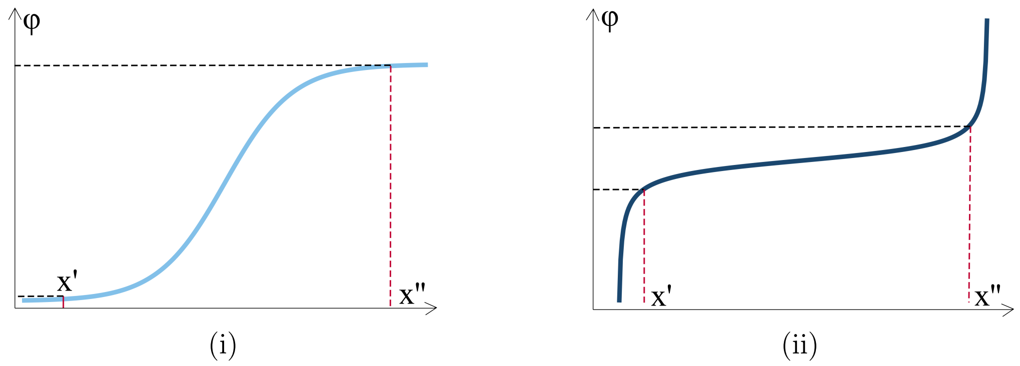


Figure 2

We should note that Corollary 3 does not state that parameter changes *always* have a larger effect on contributions when individuals face more dispersed empirical norms. If $x' > \widehat{x}$ or $x'' < \widehat{x}$, for instance, it is possible that this might not be the case. However, the result highlighted in Corollary 3 provides a rationale for why parameter shifts may have a larger effect when individuals are confronted with more dispersed descriptive norms. This is intuitive: When individuals face a dispersed descriptive norm, optimal contributions are dispersed. This implies that there is more scope for differences in parameter values to generate large swings in contributions. This intuition is corroborated by Elster and Gelfand (2021)'s cross-cultural analysis of the World Value Survey, which finds that, in loose cultures, personal values play a greater role in determining civic and pro-environmental involvement compared to tight cultures.[10]

Our final corollary moves away from individual traits and instead compares the effect of a change in the mean of the descriptive norm on contributions.

Corollary 4: *Suppose that f_2 first-order stochastically dominates f_3. Then, $x^*_{i2} \geq x^*_{i3}$ with strict inequality whenever $F_2(x^*_{i3}) < F_3(x^*_{i3})$.*

As shown in the Appendix, first-order stochastic dominance applies to all pairwise comparisons of descriptive norms with the same variance but different means in our experiment. Accordingly, Corollary 4 argues that the optimal contribution of an individual confronted with a norm exhibiting a higher mean will be higher.

2.3 Hypotheses

We can now lay out the hypotheses that follow from our theoretical framework. As shown above, there are two alternative hypotheses regarding the impact of the distribution's shape and variance, depending on the assumptions about the underlying loss function.

[10]See also d'Adda et al. (2020) who find a similar result, albeit within a context which lacks strategic uncertainty.

Hypothesis 1a. *Individuals only react to the mean of a descriptive norm, independently of its variance and shape.*

Hypothesis 1b. *Keeping everything else equal, contributions exhibit larger variance when individuals face a descriptive norm with larger variance, and tend to be polarized when individuals face a polarized descriptive norm.*

Hypothesis 1a follows directly from Proposition 1 while Hypothesis 1b follows from Proposition 2 and Corollary 1.

Consider now Corollary 2. As we will describe in greater detail below (in Section 3.3), in our experiment we elicited a measure of α_i - which we call "sucker aversion" - as well as a measure of β_i - which we call "free-riding aversion". x_i^a, an individual's personal values, are also measured directly within the experiment.

Hypothesis 2 *Suppose that Hypothesis 1 holds. Then, contributions are (i) decreasing in sucker aversion, (ii) increasing in free-riding aversion, and (iii) increasing in personal values.*

The next hypothesis is inspired by both Corollary 3 and the findings of Elster and Gelfand (2021).

Hypothesis 3. *Suppose that Hypothesis 1 holds. Keeping everything else equal, the effect of sucker aversion, free-riding aversion, and personal values on individual contributions is larger when individuals face descriptive norms with larger variance.*

Finally, Hypothesis 4 summarizes Corollary 4.

Hypothesis 4. *Keeping everything else equal, contributions are larger when individuals face descriptive norms with a higher mean.*

3 Experimental Design

3.1 Basic setup and treatment conditions

To empirically test the hypotheses derived from our theoretical framework, our experiment varies exogenously the *mean* and *variance/shape* of the co-player's behavior in a two-person PGG. Table 1 gives an overview of the different treatments, including their mean and variance. The corresponding distributions are visualized in Figure 3. Full instructions can be found in Appendix D. The experiment consists of two parts, and participants learn the details of the second part only upon completion of the first.

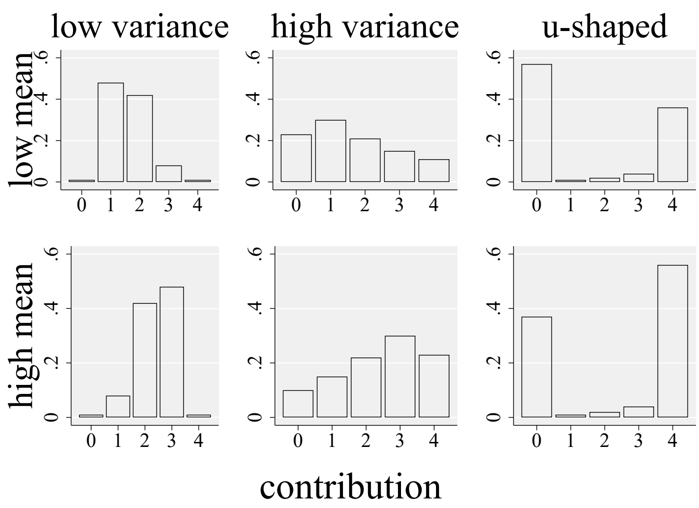
- In **Part I**, which is identical across all treatments, we use the ABC strategy method (Fischbacher et al., 2001) to elicit the underlying cooperative attitudes.
- At the beginning of **Part II**, participants are randomly assigned to one of six treatment conditions (between subjects). In each treatment, participants see a different distribution of behavior and are informed that the contribution of their co-player for Part II will be

Table 1: Experimental conditions

	Single-peaked						*Double-peaked*		
	Low variance			**High variance**			**u-shaped**		
	mean	var	mode	mean	var	mode	mean	var	mode
Low mean	1.6	0.5	3	1.6	1.6	3	1.6	3.6	4
High mean	2.4	0.5	1	2.4	1.6	1	2.4	3.6	0

randomly drawn from this distribution. The distributions vary with respect to both their mean (high and low), as well as their variance/shape (low variance, high variance, u-shaped), resulting in six treatment conditions.[11]

Figure 3: Experimental conditions: distribution of co-player's contribution



3.2 Two-player PGG and beliefs elicitation

In both Part I and Part II, we use a two-player variant of the PGG in which each participant can contribute up to four tokens (Ledyard, 1995; Fischbacher et al., 2001). Tokens invested in the public good are multiplied by 1.4 and shared between both participants. The game embodies the classic tension between private and collective interest: while fully contributing to the public good maximizes joint payoffs, each player's self-interest is maximized by contributing nothing.

To assess underlying cooperativeness, we apply the 'ABC of cooperation' (**A**ttitudes-**B**eliefs-**C**ontribution) in Part I, a method developed by Fischbacher et al. (2001) that aims to disentangle the underlying motives to contribute in a PGG. It embodies three distinct elicitations:

[11] The distributions of co-player behavior in Part II are constructed through non-random sampling from a previous session, similar to the approach adopted by e.g. by Frey and Meier (2004), Bicchieri and Xiao (2009), Krupka and Weber (2009), and Bursztyn et al. (2020). See also Charness et al. (2022) and Bardsley (2000) on the use of non-representative samples in experiments. Participants are aware that the distributions do not represent overall behavior in a PGG, but only the behavior of a selected subgroup which we constructed using real choices of subjects from a previous session. Participants understood that their behavior is incentive-compatible in that it would affect the payoffs of those previous participants, with one of which they would be paired at random.

- contribution choices conditional on each possible level of co-player's contribution to measure cooperative attitude (**A**);
- belief-elicitation task to measure expectations about co-player's contribution (**B**);
- contribution choice without being informed about co-player contribution (**C**).

We always elicit unconditional contributions first, followed by beliefs about co-player behavior and conditional contribution attitudes. The elicited attitudes give us a conditional contribution vector that we use to classify participants into different cooperation 'types'.[12] In addition, subjects are asked to express their personal values (what an individual thinks one *should* do).[13]

In line with our research question, we measure participants' beliefs as a whole distribution rather than just eliciting their beliefs about the expected co-player's contribution. To do so, we follow the approach discussed in Dimant (2022a) and ask participants to allocate points across all possible co-player contributions (see Appendix B, Figure B.2).[14] The more likely participants think a contribution is, the more points they should allocate to it. To incentivize decisions, we used a quadratic scoring rule adapted from Artinger et al. (2010), coupled with an intuitive visual interface (see Quentin, 2016) (see Appendix C for more details on the scoring rule used).

In Part II of the experiment, we ask participants to play a one-shot PGG with a randomly chosen co-player whose contribution is drawn from the shown distribution and again elicit the participants' personal values and beliefs about their co-player's contribution. Clearly enough, if the treatment is successful, then the participants' beliefs about their co-player's contribution should reflect the distribution they have been shown. The order between the different components is always randomized. The decisions participants take in the PGG have real consequences. They determine the size of the bonus for participants from the previous sessions used to construct the distributions. Figure 4 gives an overview of the design.

3.3 Sample and data collection

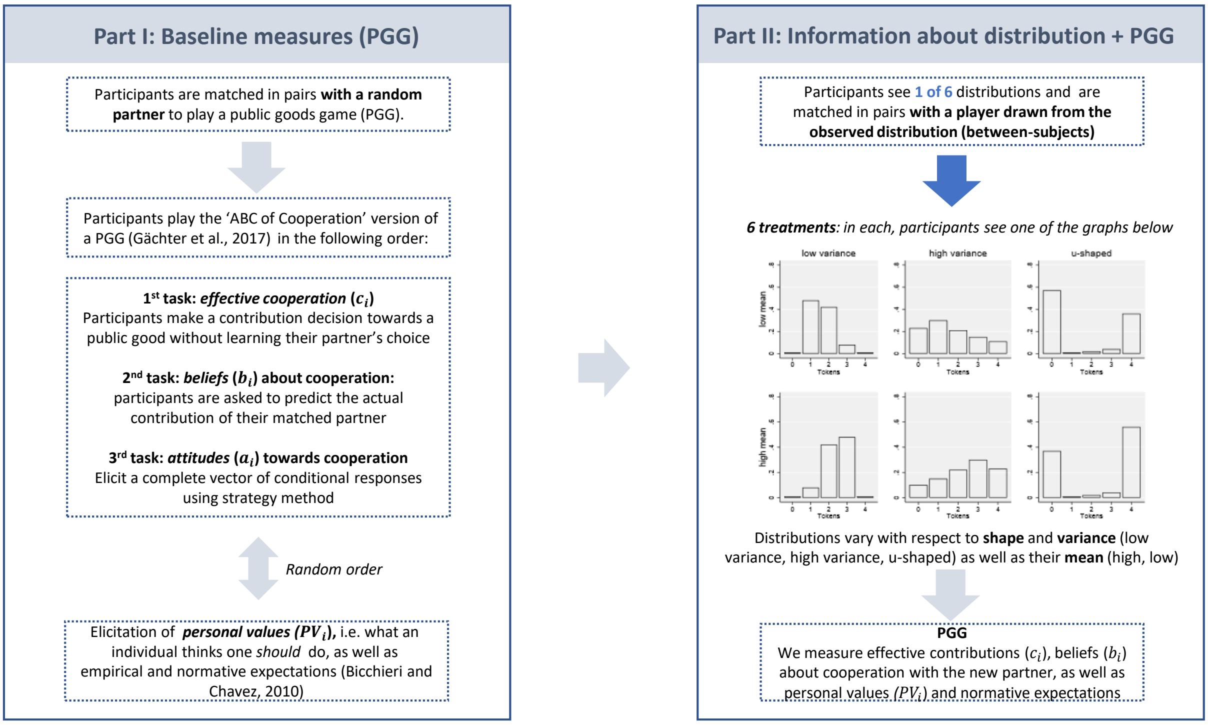
We programmed the experiment using Qualtrics (2005) and recruited participants online via Prolific in December 2021. The experiment and our hypotheses were pre-registered in November 2021. In total, we recruited a sample of about 1200 US participants that are representative in terms of age, gender, and ethnicity, resulting in about 200 observations per treatment. The chosen sample size was determined using data from a pilot and allows us to detect an effect size of $\eta^2 = 0.01$ at a 5% significance level with 90% power. On average, participants needed 17 minutes to complete the study and earned \$3.20. To construct the six distributions, we collected data from 685 MTurkers in September 2021. They initially received a show-up fee and then earned

[12] Following Fischbacher et al. (2001) we distinguish between conditional cooperators, unconditional cooperators, free-riders, triangle cooperators, and others.

[13] We also measure empirical and normative expectations about what most people do and what most people think one should do, in a randomized order. As shown in Appendix B, (Figure B.5), these are highly correlated with personal values and beliefs about the behavior of the co-player.

[14] This method is inspired by existing work on eliciting distributions of subjective beliefs (e.g., Lau et al., 1998; Goldstein and Rothschild, 2014; Harrison et al., 2017) as well as an oral presentation given by Don Ross at the 2019 "Norms and Behavior Change" (NoBeC) workshop. For the purposes of our investigation, we apply those insights to our context of tight, loose, and polarized environments.

Figure 4: Overview of the experimental design



an additional bonus depending on the decisions of the participants in the main experiment. Participants are paid for both their decision in Parts I and II, but receive no information about their payoffs between parts to reduce potential hedging.

After completing the experiment, participants completed an ex post survey that provided additional demographic controls. In addition, we asked participants about the perceived average and variance of the observed distributions as well as their difficulty in interpreting them. As we told participants that the distribution from which their co-player's contribution is drawn was taken from one of the six subgroups we constructed from previous sessions, we also asked how common they think this behavior is.[15] Finally, we introduced two questions to proxy the participants' aversion towards contributing more and less than their co-player (sucker aversion and free-riding aversion).

4 Results

4.1 Behavior in Part I

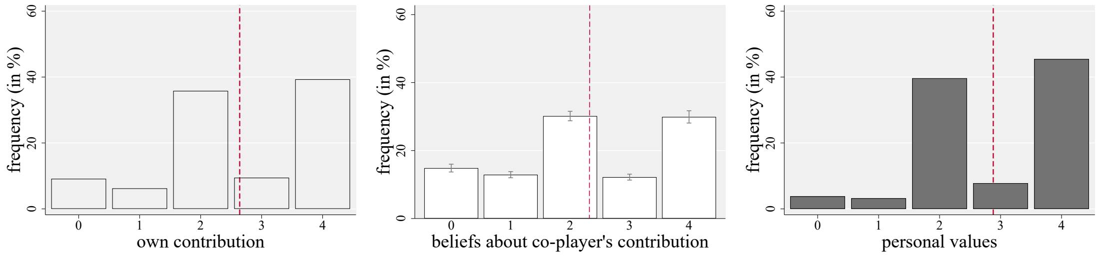
First, we provide an overview of our Part I results. Consistent with the existing literature, the contribution schedule in our data reveals a strong pattern of conditional cooperation among

[15] Figure B.6 in Appendix B shows that participants have a correct interpretation of the observed distributions and state that it was relatively easy to understand them. Moreover, the perception of how difficult to interpret and how common the observed behavior is in a wider population is similar across conditions. On a scale from 1 (very rare) to 7 (very common), participants gave a rating of 5 to the distribution they saw, suggesting they believed it to be fairly common.

participants. Following the type definitions developed by Fischbacher et al. (2001) and Thöni and Volk (2018), we classify 84% of all participants as *conditional cooperators*.[16] The distribution of types is independent of our treatments (χ^2-test, $p = 0.35$).

Figure 5 shows contributions, aggregate beliefs about co-player contributions, and personal values in Part I. The most frequent contribution levels are 2 and 4 tokens, revealing relatively high levels of cooperation. The same is true for beliefs about the other player's contribution and personal values about what one should do in the game.[17]

Figure 5: Contributions, aggregate beliefs about the co-player's contribution, and personal values in Part I



Note. The dashed lines represent averages. Whiskers show 95% confidence intervals.

The data from Part I already allow us to gain some insight into our research question. Although we do not have exogenous variation in co-player behavior, we can look at the relationship between participants' contributions and the variance of their beliefs about their co-player's contribution. We find that for participants who show a greater variation in beliefs, contributions also vary more (F-test, $p = 0.001$). While this analysis cannot provide causal evidence, it gives initial anecdotal support for the notion that variance of one's co-player's behavior matters and that looser environments may generate more dispersed responses.[18] In the next section, we turn to a more rigorous test of our hypotheses that builds on exogenous variation in co-player's behavior.

4.2 Effect of variance and shape of the descriptive norm

This section addresses how *exogenous* differences in the variance and shape of the descriptive norm presented to participants affect individual responses. We start by looking at the participants' beliefs about their co-player's contribution. Figure 6 shows that these beliefs closely mirror the distribution of co-player behavior that was presented to participants.[19]

Next, we turn to our main research question of whether differences in the variance and shape of norms affect individual contribution behavior. Figure 7 shows the distribution of contributions in Part II for each experimental condition, confirming that there is indeed a stark difference between treatments. In particular, we see that in tight environments (low variance), participants

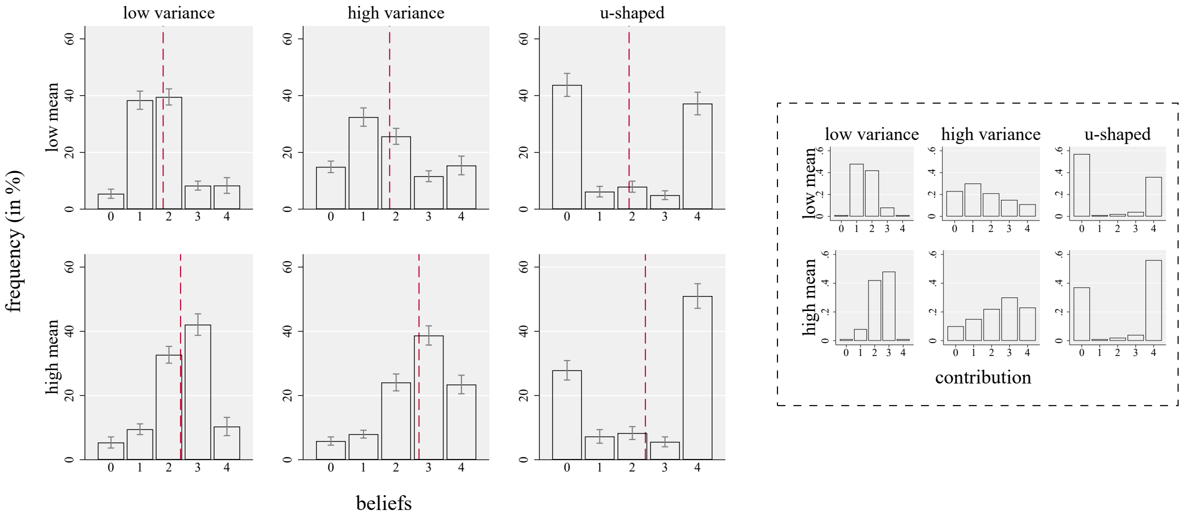
[16] The rest consists of 5% unconditional cooperators, 3% free-riders, 5% triangle cooperators and 4% others.

[17] In Appendix B (Figures B.3 and B.4) we provide a more detailed characterization of the participants' beliefs about their co-player's contribution in Part I, showing that the aggregate distribution in Figure 5 hides substantial heterogeneity.

[18] Similarly, we find a significant positive correlation between average beliefs and contributions (r=0.65, $p < 0.001$).

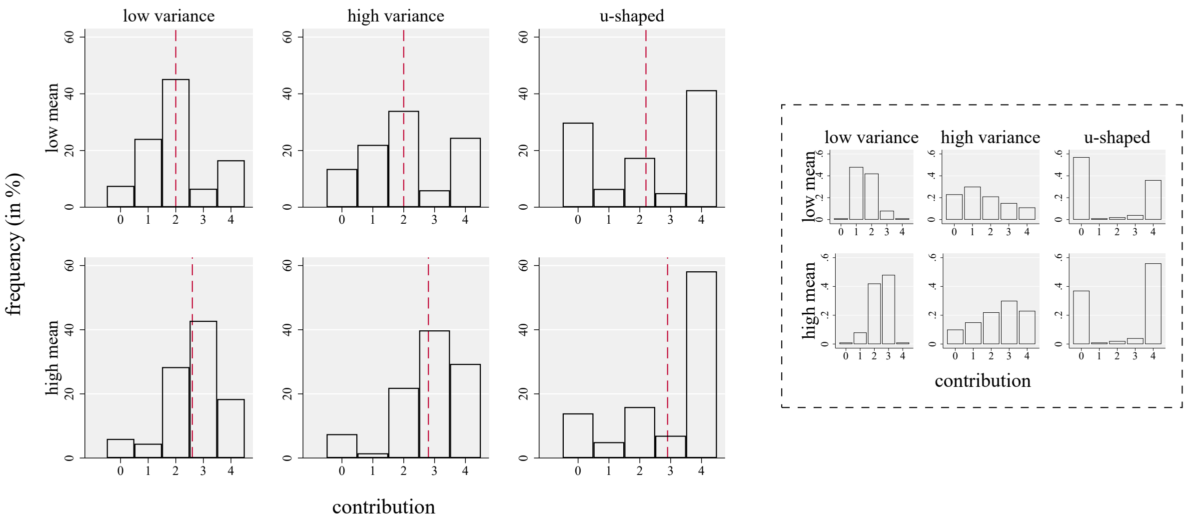
[19] Figure B.10 in Appendix B visualises how participants change their beliefs between parts as a reaction to the provided descriptive norm.

Figure 6: Aggregate beliefs about co-player's contribution in Part II by treatment



Note. The dashed lines represent average beliefs about co-player's contribution. Whiskers show 95% confidence intervals. The righthand box depicts the distributions shown in each treatment.

Figure 7: Distribution of participants' contributions in Part II by treatment



Note. The dashed lines represent average contributions. The righthand box depicts the distributions shown in each treatment.

choose contribution levels that are tightly centered around the mean of the shown distribution ($\sigma^2 = 1.26$). In loose environments (high variance), by contrast, we see a much larger variation in behavior ($\sigma^2 = 1.65$). In other words, loose behavior generates loose responses while tight behavior generates tight responses. We also find that in polarized environments, participants show very heterogeneous reactions ($\sigma^2 = 2.68$). In Section 4.3 we provide a discussion of the personal traits that drive this heterogeneity.

To test the first visual impression of treatment effects, we perform pairwise F-tests for the equality of standard deviations between treatments. Overall, we find that the variance in contributions is significantly higher in polarized environments than in tight or loose environments (for both $F < 0.001$).[20] Loose conditions in turn generate a significantly higher variance in participants' contributions than tight conditions ($F = 0.007$). Pairwise χ^2 tests also confirm that the distribution of contributions is significantly different between treatments ($p < 0.001$, $p < 0.001$, and $p = 0.002$ respectively). Moreover, the distribution of contributions in Part II is significantly different from the distribution in Part I ($p < 0.001$). As a further test of treatment differences, in Appendix B we split the sample into two groups, high contributions (> 2) and low contributions (≤ 2), and show that a higher variance in co-player behavior *increases* high contributions, while it *decreases* low ones (see Table B.1). This confirms that the variance of behavior increases when participants are faced with more spread-out descriptive norms. Finally, Figure 7 shows that the polarized norm induces polarization in subsequent contribution behavior.

Taken together, our results confirm the importance of both the variance and shape of the observed behavior for individual decisions. Therefore, we reject Hypothesis 1a and accept Hypothesis 1b. Different environments generate very different responses. Loose, tight, and polarized behaviors reproduce themselves.

Result 1. *Looser descriptive norms lead to a larger variance in contributions. Polarized descriptive norms generate extreme contributions.*

4.3 Personal traits

We now look at the effect of personal values, free-riding, and sucker aversion on contributions (see Table 2).[21] To account for the censored nature of our data (Tobin, 1958) and in line with previous PGG studies (see e.g. Fehr and Gächter, 2000; Fischbacher and Gächter, 2006; Chaudhuri et al., 2017), we use Tobit regressions to analyze contributions.[22] Consistent with Hypothesis 2, Table 2 confirms that personal values and free-riding aversion increase contributions, while sucker aversion decreases them.

Result 2. *Sucker-aversion lowers contributions while free-riding aversion and personal values increase contributions.*

Consider now Hypothesis 3: Personal traits should matter more in loose or polarized environments compared to tight ones. Intuitively, when strategic uncertainty is high, individuals face a trade-off. If they increase their contribution, they face a higher probability of looking like a sucker, by contributing more than their co-player. On the other hand, by decreasing their contribution they face a higher chance of looking like a free-rider, by contributing less than their co-player. Their degree of sucker aversion and free-riding aversion, as well as their personal values, determine the outcome of this trade-off.

As can be seen from Table 2, personal traits guide individual contributions more in environments characterized by high strategic uncertainty. The interactions between the u-shaped

[20] Figure B.7 in Appendix B shows the distribution of contributions across different variance conditions, independent of the mean. Figure B.8 shows that the result holds when restricting the sample to conditional cooperators.

[21] Table B.2 in Appendix B shows the same analysis focusing on the sub-sample of conditional cooperators.

[22] Alternative estimation methods such as OLS or ordered probits yield qualitatively similar results.

environment and sucker aversion, as well as between u-shaped environment and free-riding aversion are highly significant in the expected direction. Personal values have a significantly stronger effect on individual contributions in both high-variance and u-shaped environments (as can be seen by looking at the interaction between personal values and high variance and between personal values and u-shaped environment in Table 2).[23] This supports Hypothesis 3.

Result 3. *Personal traits (sucker-version, free-riding aversion, and personal values) matter more when participants are confronted with loose/polarized descriptive norms.*

The finding that personal traits moderate participants' reaction to environments characterized by high strategic uncertainty explains the greater heterogeneity observed in those environments relative to tight environments, where strategic uncertainty is very low. Applied to the polarized scenarios, this means that whether a participant's contribution appears to respond primarily to (potential) co-player contributions at the lower or higher end of the distribution seems to be at

Table 2: Tobit models. Effect of personal values, sucker, and free-riding aversion on contributions in Part II

	(1)	(2)
Sucker aversion	-0.16***	-0.16***
	(0.06)	(0.06)
Sucker aversion x high variance	-0.04	-0.00
	(0.08)	(0.08)
Sucker aversion x u-shaped	-0.26***	-0.24***
	(0.08)	(0.08)
Free-riding aversion	0.17***	0.17***
	(0.05)	(0.05)
Free-riding aversion x high variance	0.03	0.01
	(0.07)	(0.07)
Free-riding aversion x u-shaped	0.27***	0.27***
	(0.07)	(0.07)
Personal values (PVs)	0.52***	0.50***
	(0.09)	(0.09)
PVs x high variance	0.29**	0.29**
	(0.12)	(0.12)
PVs x u-shaped	0.37***	0.38***
	(0.13)	(0.13)
Constant	0.65*	-1.45**
	(0.38)	(0.70)
Demographic controls	No	Yes
N observations	1203	1188
Pseudo R^2	0.10	0.12

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.
Note. Regressions control for mean and variance of the observed distribution. *Personal values* are measured in Part I and can take values between 0 and 4. Sucker and free-riding aversion are measured on a Likert scale from 1 to 7. Demographic controls include age, gender, education, acceptance of risk, trust, and measures for negative and positive reciprocity. All regressions control for order effects.

[23] We use personal values in Part I as these are collected "in a vacuum" and are therefore not affected by the treatments. Our results hold when controlling for the change in personal values and their interaction with treatment indicators (see Appendix B Table B.3).

least partly driven by the participant's personal traits.[24]

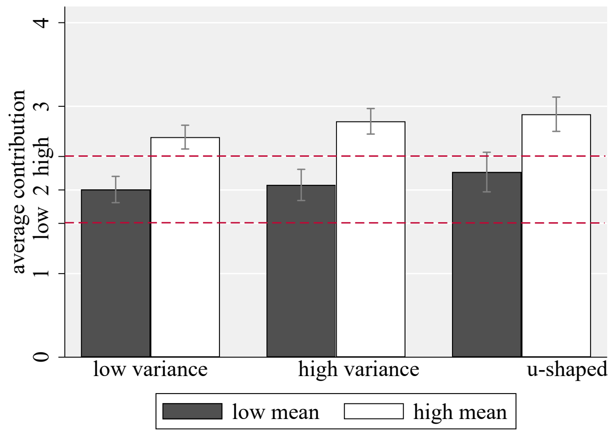
4.4 Effect of high and low means

Finally, we look at the difference between descriptive norms with high and low means. In line with previous literature, contributions are significantly higher in high mean conditions (Wilcoxon-Mann-Whitney test, $p < 0.001$). This is true independent of the shape and variance (Figure 8).

Table 3 tests Hypothesis 4 formally. As expected, contributions are significantly higher in the high mean conditions. This finding also holds when controlling for baseline behavior and including demographic controls (see Models 2 and 3). Moreover, models 1-3 show that, overall, contributions are higher in the high-variance and polarized conditions than in the low-variance conditions. This is an interesting finding worth emphasizing: In loose and polarized environments, participants could in principle use the non-negligible probability of facing a free-rider as an "alibi" to justify self-serving selfish behavior. However, our data suggest that this is not the case. Greater strategic uncertainty promotes higher contributions on average. Following the findings discussed in Result 2, this appears to be driven by many participants experiencing high levels of free-riding aversion, as well as high personal values for cooperation. It would be thus interesting to investigate this relationship in other contexts where personal values may diverge from our sample. Finally, Table 3 shows that the effect of the mean seems to be even larger in the polarized condition, as indicated by the significant interaction (see models 4-6).

Result 4. *Participants contribute significantly more when the descriptive norm has a higher mean.*

Figure 8: Effect of a high or low mean on contributions



Note. The dashed lines represent the mean of the observed distributions (high = 2.4, low = 1.6). Whiskers show 95% confidence intervals.

[24]In Appendix B, Table B.4, we look at the correlation between personal traits and demographic characteristics. Women score higher than men both in terms of free-riding and sucker aversion, while older participants have lower scores in both.

Table 3: Tobit models. Effect of high and low mean conditions on contributions in Part II

	No interaction			Interaction		
	(1)	(2)	(3)	(4)	(5)	(6)
High mean	1.00***	0.99***	0.93***	0.70***	0.64***	0.58***
	(0.14)	(0.11)	(0.10)	(0.23)	(0.18)	(0.17)
Variance (baseline = low)						
High variance	0.23	0.26**	0.26**	0.08	0.09	0.09
	(0.16)	(0.13)	(0.12)	(0.23)	(0.18)	(0.17)
U-shaped	0.61***	0.66***	0.61***	0.29	0.28	0.24
	(0.17)	(0.13)	(0.13)	(0.23)	(0.18)	(0.18)
Interactions						
High mean x high variance				0.29	0.33	0.35
				(0.33)	(0.26)	(0.25)
High mean x u-shaped				0.65*	0.80***	0.78***
				(0.34)	(0.27)	(0.26)
Constant	2.05***	-1.10***	-1.77***	2.21***	-0.93***	-1.56***
	(0.16)	(0.21)	(0.58)	(0.19)	(0.22)	(0.58)
Baseline controls	No	Yes	Yes	No	Yes	Yes
Demographic controls	No	No	Yes	No	No	Yes
N observations	1203	1203	1188	1203	1203	1188
Pseudo R^2	0.02	0.14	0.17	0.02	0.14	0.17

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

Note. Data is censored at 0 and 4. *High mean* is a binary variable with 0 = low and 1 = high mean. *Variance* is a categorical variable with 0 = low variance, 1 = high variance and 2 = u-shaped. Baseline controls include contributions, beliefs, and personal values in Part I and can take values between 0 and 4. Demographic controls include age, gender, education, acceptance of risk, trust, sucker aversion, free-riding aversion, and measures for negative and positive reciprocity. All regressions control for order effects.

5 Discussion and conclusion

In this study, we investigate how different descriptive norms of cooperative behavior affect an individual's own willingness to cooperate. Our contribution lies in the causal evaluation of the tightness-looseness framework of norms that has its origin in cross-cultural psychology (Gelfand et al., 2011) through the lens of an economic model that we test with a behavioral experiment. We first develop a theoretical framework that is based on the notion that individuals are motivated by reciprocity concerns and that differences in the variance/shape of descriptive norms generate different degrees of strategic uncertainty, which in turn affect individual behavior. We then test our framework empirically in the context of a public goods game where we vary both the *mean* (high/low) as well as the *variance* and *shape* (tight/loose/polarized) of the distribution from which the co-player's contribution is drawn.

Our results confirm previous research showing that information about average behavior has an important effect on subsequent decisions. Individuals contribute significantly more in high mean conditions than in low mean conditions. Most importantly, though, we show that the mean is not the only relevant feature of the distribution. In line with our theoretical framework, we find that loose norms generate a larger variance in individual responses compared to tight norms, and that polarized environments generate polarized behavior. In other words, "tight breeds tight", "loose breeds loose", and "polarized breeds polarized". Interestingly, we find that a large share of participants makes choices that expose them to considerable strategic risk. Consider, for instance, a situation where a share $p < 1/2$ of the population contributes 4 and a share $1 - p$

contributes 0. This is an approximation of the descriptive norm that we utilized in our polarized treatment with a low mean. In this context, choosing a contribution of 2 allows eliminating all strategic risk, since it generates a *sure* mismatch of 2 between one's own contribution and the co-player's. Anyone who is risk-averse in terms of mismatches will therefore not choose to contribute more than 2, as this exposes them to avoidable strategic risk, while also lowering material payoff.[25] Yet, we find that a sizeable share of our subjects (46%, of which 41% contribute 4) does precisely that. Although (as indicated in Part I of the experiment) the overwhelming majority of participants are strongly motivated by reciprocity concerns and thus want to match their co-player's contribution, they are not mismatch-risk-averse. This suggests that, when it comes to the social domain, individuals may exhibit different attitudes to risk compared to what we are accustomed to seeing in the monetary domain, which is typically characterized by risk-averse behavior.[26] In future research, one could investigate this hypothesis further to fully appreciate the potential discrepancy between attitudes toward strategic risk and attitudes toward risk in the monetary domain.

Another key finding of our analysis is that an individual's reaction to high strategic uncertainty is moderated by their personal traits. When faced with high strategic uncertainty, people's responses are heterogeneous. Relative to environments where strategic uncertainty is low, decisions under high uncertainty are more strongly influenced by personal traits such as personal values. This has practical implications for policymakers and behavioral interventions. Current interventions are often directed at both personal values and beliefs, as well as norms to achieve change (Dimant and Shalvi, 2022). Our results suggest that depending on the relative tightness or looseness of the norm, different approaches might be more fruitful. For example, when intervening in contexts with loose or polarized norms, a focus on personal values might be more successful, whereas when intervening in contexts with tight norms, it may be better to focus on the behaviors of others.

As is universally true for experimental research, the existence and role of experimenter demand effects (EDEs) should be considered (Zizzo, 2010). We consider the potential presence of EDEs to not be a concern in our setup for two reasons. First, the finding that loose norms lead to loose responses and polarized norms to polarized ones does not rely on the behavior of one individual but results from the behavior of all participants. Thus, different people respond very differently to the observed distribution. This makes demand effects with respect to variance improbable. Second, demand effects would not be able to explain the interaction between variance and personal values that we observe in our results.

Our analysis provides many other avenues for future research. For example, existing literature suggests that the enforcement of norms through punishment is an important part of sustaining existing norms (Balafoutas and Nikiforakis, 2012; Balafoutas et al., 2014). To test our research question, the inclusion of a punishment opportunity was not essential. Our results show that

[25] For instance, choosing to contribute 4 exposes a participant to a lottery where with probability $1-p$ the mismatch between their own and the co-player's contribution is 4, and with probability p it is 0. It is easy to see that $v(2) < (1-p)v(4) + pv(0)$ for any increasing convex loss function $v(.)$. This follows since, by Jensen's inequality, $(1-p)v(4) + pv(0) > v(4(1-p)) > v(2)$ where the last inequality is due to $p < 1/2$.

[26] See e.g. Chetty (2006) and more generally the vast literature on monetary risk aversion.

tight/loose/polarized descriptive norms generate different responses even in the *absence* of a norm enforcement mechanism. Future research could investigate the role of these behavioral patterns in a setup that incorporates punishment. Just as we have shown that individuals are sensitive to distributions of contributions, they are also likely to be responsive to distributions of punishment and adjust their behavior accordingly. Another potential extension of our work would be to move beyond WEIRD (Western, Educated, Industrialized, Rich, Democratic, see Henrich et al., 2010) samples to test the generalizability of our results.

Although the analysis is static, we believe that our results can speak to the long-run sustainability of different descriptive norms (tight/loose/polarized). A necessary condition for a norm to be self-sustaining (in the sense of the standard definition of a stationary distribution (see e.g. Ross et al., 1996) is that, when individuals are confronted with the norm, their reactions should reproduce the norm itself. This is very much in line with our findings. Our participants strongly react to the shape of the descriptive norm they are presented with and the resulting distributions look remarkably similar to the original norms. In this sense, our results can be seen as providing suggestive evidence that – all else equal – there may be multiple equilibria, involving tight, loose, or polarized distributions of behavior. This ties our findings to the (primarily theoretical) literature on norms as equilibrium selection devices (see e.g. Binmore and Samuelson, 1994; Basu, 1998; Young, 2015) and also opens the door for future research on exploring tightness and looseness in a dynamic and fully endogenous setting.

Overall, we show that considering the whole distribution instead of focusing only on average behavior provides substantial analytical richness. This can form the basis for a better appreciation of different behavioral patterns observed across societies. We hope that our work will pave the way to a wider understanding of the interplay between norms and behavior that encompasses less-studied aspects such as variance and shape, generating a fertile agenda for future research.

References

Adriani, F. and Sonderegger, S. (2019). A theory of esteem based peer pressure. *Games and Economic Behavior*, 115:314–335.

Artinger, F., Exadaktylos, F., Koppel, H., and Sääksvuori, L. (2010). Applying quadratic scoring rule transparently in multiple choice settings: A note. Technical report, *Jena Economic Research Papers*.

Balafoutas, L. and Nikiforakis, N. (2012). Norm enforcement in the city: A natural field experiment. *European Economic Review*, 56(8):1773–1785.

Balafoutas, L., Nikiforakis, N., and Rockenbach, B. (2014). Direct and indirect punishment among strangers in the field. *Proceedings of the National Academy of Sciences*, 111(45):15924–15927.

Bardsley, N. (2000). Control without deception: Individual behaviour in free-riding experiments revisited. *Experimental Economics*, 3(3):215–240.

Barr, A., Dekker, M., Mwansa, F., and Zuze, T. L. (2020). Financial decision-making, gender and social norms in zambia: Preliminary report on the quantitative data generation, analysis and results. Technical report, CeDEx Discussion Paper Series.

Bašić, Z. and Verrina, E. (2020). Personal norms—and not only social norms—shape economic behavior. *MPI Collective Goods Discussion Paper,(2020/25)*.

Basu, K. (1998). Social norms and the law. *The New Palgrave Dictionary of Economics and the Law,(Peter Newman, ed.)*.

Bauer, K., Chen, Y., Hett, F., and Kosfeld, M. (2022). Group identity and opinion formation: With applications to political polarization. Mimeo.

Bicchieri, C. (2005). *The grammar of society: The nature and dynamics of social norms*. Cambridge University Press.

Bicchieri, C. and Chavez, A. (2010). Behaving as expected: Public information and fairness norms. *Journal of Behavioral Decision Making*, 23(2):161–178.

Bicchieri, C. and Dimant, E. (2019). Nudging with care: The risks and benefits of social information. *Public Choice*, pages 1–22.

Bicchieri, C., Dimant, E., Gächter, S., and Nosenzo, D. (2022a). Social proximity and the erosion of norm compliance. *Games and Economic Behavior*, 132:59–72.

Bicchieri, C., Dimant, E., Gelfand, M., and Sonderegger, S. (2022b). Social norms and behavior change: The interdisciplinary research frontier. *Journal of Economic Behavior & Organization*.

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

Binmore, K. and Samuelson, L. (1994). An economist's perspective on the evolution of norms. *Journal of Institutional and Theoretical Economics (JITE)/Zeitschrift für die gesamte Staatswissenschaft*, pages 45–63.

Bowles, S. and Gintis, H. (2013). *A Cooperative Species: Human reciprocity and its evolution*. Princeton University Press.

Bursztyn, L., Egorov, G., and Fiorin, S. (2020). From extreme to mainstream: The erosion of social norms. *American Economic Review*, 110(11):3522–48.

Callander, S. and Carbajal, J. C. (2022). Cause and effect in political polarization: A dynamic analysis. *Journal of Political Economy*, 130(4):825–880.

Capraro, V., Jagfeld, G., Klein, R., Mul, M., and de Pol, I. v. (2019). Increasing altruistic and cooperative behaviour with simple moral nudges. *Scientific Reports*, 9(1):1–11.

Charness, G., Samek, A., and van de Ven, J. (2022). What is considered deception in experimental economics? *Experimental Economics*, 25(2):385–412.

Chaudhuri, A., Paichayontvijit, T., et al. (2006). Conditional cooperation and voluntary contributions to a public good. *Economics Bulletin*, 3(8):1–14.

Chaudhuri, A., Paichayontvijit, T., and Smith, A. (2017). Belief heterogeneity and contributions decay among conditional cooperators in public goods games. *Journal of Economic Psychology*, 58:15–30.

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

Chetty, R. (2006). A new method of estimating risk aversion. *American Economic Review*, 96(5):1821–1834.

Cheung, S. L. (2014). New insights into conditional cooperation and punishment from a strategy method experiment. *Experimental Economics*, 17(1):129–153.

Cialdini, R. B., Reno, R. R., and Kallgren, C. A. (1990). A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology*, 58(6):1015.

Cialdini, R. B. and Trost, M. R. (1998). Social influence: Social norms, conformity and compliance. In Gilbert, D. T., Fiske, S. T., and Lindzey, G., editors, *The Handbook of Social Psychology*, pages 151–192. New York: McGraw-Hill.

Croson, R. T. (2007). Theories of commitment, altruism and reciprocity: Evidence from linear public goods games. *Economic Inquiry*, 45(2):199–216.

d'Adda, G., Dufwenberg, M., Passarelli, F., and Tabellini, G. (2020). Social norms with private values: Theory and experiments. *Games and Economic Behavior*, 124:288–304.

DeGroot, M. H. (2005). *Optimal statistical decisions.* John Wiley & Sons.

Dimant, E. (2019). Contagion of pro-and anti-social behavior among peers and the role of social proximity. *Journal of Economic Psychology*, 73:66–88.

Dimant, E. (2022a). Distributions matter: Measuring the tightness and looseness of social norms. Working Paper Available at SSRN: https://dx.doi.org/10.2139/ssrn.4107802.

Dimant, E. (2022b). Hate trumps love: The impact of political polarization on social preferences. *Management Science.*

Dimant, E., Clemente, E. G., Pieper, D., Dreber, A., and Gelfand, M. (2022). Politicizing mask-wearing: predicting the success of behavioral interventions among republicans and democrats in the us. *Scientific reports*, 12(1):1–12.

Dimant, E. and Shalvi, S. (2022). Meta-nudging honesty: Past, present, and future of the research frontier. *Current Opinion in Psychology*, page 101426.

Elster, A. and Gelfand, M. J. (2021). When guiding principles do not guide: The moderating effects of cultural tightness on value-behavior links. *Journal of Personality*, 89(2):325–337.

Enke, B. (2020). Moral values and voting. *Journal of Political Economy*, 128(10):3679–3729.

Fehr, E. and Gächter, S. (2000). Fairness and retaliation: The economics of reciprocity. *Journal of Economic Perspectives*, 14(3):159–181.

Fehr, E. and Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, 114(3):817–868.

Fehr, E. and Schurtenberger, I. (2018). Normative foundations of human cooperation. *Nature Human Behaviour*, 2(7):458–468.

Feldhaus, C., Sobotta, T., and Werner, P. (2019). Norm uncertainty and voluntary payments in the field. *Management Science*, 65(4):1855–1866.

Fischbacher, U. and Gächter, S. (2006). Heterogeneous social preferences and the dynamics of free riding in public goods. *IZA discussion paper No.2011.*

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

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

Fisman, R. and Miguel, E. (2007). Corruption, norms, and legal enforcement: Evidence from diplomatic parking tickets. *Journal of Political economy*, 115(6):1020–1048.

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

Gächter, S., Herrmann, B., and Thöni, C. (2010). Culture and cooperation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1553):2651–2661.

Gächter, S., Kölle, F., and Quercia, S. (2017). Reciprocity and the tragedies of maintaining and providing the commons. *Nature Human Behaviour*, 1(9):650–656.

Gelfand, M. J. (2021). Cultural evolutionary adaptations to threat. Current directions in psychological science. In press.

Gelfand, M. J., Raver, J. L., Nishii, L., Leslie, L. M., Lun, J., Lim, B. C., Duan, L., Almaliach, A., Ang, S., Arnadottir, J., et al. (2011). Differences between tight and loose cultures: A 33-nation study. *Science*, 332(6033):1100–1104.

Goldstein, D. G. and Rothschild, D. (2014). Lay understanding of probability distributions. *Judgment & Decision Making*, 9(1).

Grout, P. A., Mitraille, S., and Sonderegger, S. (2015). The costs and benefits of coordinating with a different group. *Journal of Economic Theory*, 160:517–535.

Harrison, G. W., Martínez-Correa, J., Swarthout, J. T., and Ulm, E. R. (2017). Scoring rules for subjective probability distributions. *Journal of Economic Behavior & Organization*, 134:430–448.

Hartig, B., Irlenbusch, B., and Kölle, F. (2015). Conditioning on what? heterogeneous contributions and conditional cooperation. *Journal of Behavioral and Experimental Economics*, 55:48–64.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., and McElreath, R. (2001a). In search of homo economicus: behavioral experiments in 15 small-scale societies. *American Economic Review*, 91(2):73–78.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., McElreath, R., et al. (2001b). Cooperation, reciprocity and punishment in fifteen small-scale societies. *American Economic Review*, 91(2):73–78.

Henrich, J., Heine, S. J., and Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, 33(2-3):61–83.

Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., and Westwood, S. J. (2019). The origins and consequences of affective polarization in the united states. *Annual Review of Political Science*, 22:129–146.

Iyengar, S. and Westwood, S. J. (2015). Fear and loathing across party lines: New evidence on group polarization. *American Journal of Political Science*, 59(3):690–707.

Kandel, E. and Lazear, E. P. (1992). Peer pressure and partnerships. *Journal of Political Economy*, 100(4):801–817.

Kerr, N. L., Rumble, A. C., Park, E. S., Ouwerkerk, J. W., Parks, C. D., Gallucci, M., and Van Lange, P. A. (2009). "how many bad apples does it take to spoil the whole barrel?": Social exclusion and toleration for bad apples. *Journal of Experimental Social Psychology*, 45(4):603–613.

Kölle, F. and Quercia, S. (2021). The influence of empirical and normative expectations on cooperation. *Journal of Economic Behavior & Organization*, 190:691–703.

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

Krupka, E. L. and Weber, R. A. (2013). Identifying social norms using coordination games: Why does dictator game sharing vary? *Journal of the European Economic Association*, 11(3):495–524.

Lau, H.-S., Lau, A. H.-L., and Ho, C.-J. (1998). Improved moment-estimation formulas using more than three subjective fractiles. *Management Science*, 44(3):346–351.

Ledyard, J. O. (1995). Public goods: A survey of experimental research. In Roth, A. and Kagel, J., editors, *Handbookof Experimental Economics*. Princeton University Press.

Machina, M. and Pratt, J. (1997). Increasing risk: some direct constructions. *Journal of Risk and Uncertainty*, 14(2):103–127.

McConnell, C., Margalit, Y., Malhotra, N., and Levendusky, M. (2018). The economic consequences of partisanship in a polarized era. *American Journal of Political Science*, 62(1):5–18.

Michaeli, M. and Spiro, D. (2015). Norm conformity across societies. *Journal of public economics*, 132:51–65.

Michaeli, M. and Spiro, D. (2017). From peer pressure to biased norms. *American Economic Journal: Microeconomics*, 9(1):152–216.

Murphy, A. H. and Winkler, R. L. (1970). Scoring rules in probability assessment and evaluation. *Acta Psychologica*, 34:273–286.

Nunn, N. (2022). On the dynamics of human behavior: The past, present, and future of culture, conflict, and cooperation. In *AEA Papers and Proceedings*, volume 112, pages 15–37.

Perez-Truglia, R. and Cruces, G. (2017). Partisan interactions: Evidence from a field experiment in the united states. *Journal of Political Economy*, 125(4):1208–1243.

Qualtrics (2005). Qualtrics software, Version [January 2021] of Qualtrics. Copyright ©[2021]. Qualtrics, Provo, Utha, USA. Available at: https://www.qualtrics.com.

Quentin, A. (2016). distBuilder v 1.1.0 (1.1.0). Zenodo. https://doi.org/10.5281/zenodo.166736.

Ren, Z. B., Dimant, E., and Schweitzer, M. (2023). Beyond belief: How social engagement motives influence the spread of conspiracy theories. *Journal of Experimental Social Psychology*, 104:104421.

Robbett, A. and Matthews, P. H. (2022). Polarization and group cooperation. *Political Behavior*.

Ross, S. M., Kelly, J. J., Sullivan, R. J., Perry, W. J., Mercer, D., Davis, R. M., Washburn, T. D., Sager, E. V., Boyce, J. B., and Bristow, V. L. (1996). *Stochastic Processes*, volume 2. Wiley New York.

Shang, J. and Croson, R. (2009). A field experiment in charitable contribution: The impact of social information on the voluntary provision of public goods. *The Economic Journal*, 119(540):1422–1439.

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

Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica: Journal of the Econometric Society*, pages 24–36.

Winkler, M. (2021). Do disasters affect the tightness of social norms. *Working Paper*.

Wolff, I. (2017). What are the equilibria in public-good experiments? *Economics Letters*, 150:83–85.

Young, H. P. (2015). The evolution of social norms. *Annual Review of Economics*, 7(1):359–387.

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

A Theoretical Appendix

Proof of proposition 1 Expected payoff is

$$X - x_i + \gamma (x_i + x_j) - \frac{\eta_i}{2} (x_i - x_j)^2 - \frac{\delta_i}{2} (x_i - x_i^a)^2$$

which can be rewritten as

$$X - x_i + \gamma (x_i + x_j) - \frac{\eta_i}{2} \int_0^{\overline{x}} (x_i - x_j)^2 f(x_j) dx_j - \frac{\delta_i}{2} (x_i - x_i^a)^2 . \tag{3}$$

The derivative of (3) with respect to x_i gives

$$-(1-\gamma) - \eta_i (x_i - \mu) - \delta_i (x_i - x_i^a) . \tag{4}$$

Evaluated at $x_i = 0$, (4) becomes $-(1-\gamma) + \eta_i \mu + \delta_i x_i^a$. When $\eta_i < \frac{1-\gamma-\delta_i x_i^a}{\mu}$ we therefore have $x_i^* = 0$. Evaluated at $x_i = \overline{x}$, (4) becomes $-(1-\gamma) - \eta_i (\overline{x} - \mu_x) - \delta_i (\overline{x} - x_i^a) < 0$. Finally, as highlighted in the proposition, in any interior solution (4) is equal to 0, so that $x_i^* = \frac{\eta_i \mu + \delta_i x_i^a - (1-\gamma)}{\eta_i + \delta_i}$. ■

Proof of proposition 2 Expected payoff is:

$$\begin{aligned} & X - x_i + \gamma (x_i + x_j) - \alpha_i F(x_i) [x_i - E(x_j \mid x_j < x_i)] \\ & -\beta_i (1 - F(x_i)) [E(x_j \mid x_j > x_i) - x_i] - \frac{\delta_i}{2} (x_i - x_i^a)^2 \end{aligned}$$

which can be rewritten as

$$\begin{aligned} & X - x_i + \gamma (x_i + x_j) - \alpha_i \int_0^{x_i} (x_i - x_j) f(x_j) dx_j \\ & -\beta_i \int_{x_i}^{\overline{x}} (x_j - x_i) f(x_j) dx_j - \frac{\delta_i}{2} (x_i - x_i^a)^2 . \end{aligned} \tag{5}$$

The derivative of (5) with respect to x_i gives

$$-(1-\gamma) - \alpha_i F(x_i) + \beta_i (1 - F(x_i)) - \delta_i (x_i - x_i^a) . \tag{6}$$

Evaluated at $x_i = 0$, (6) becomes $\beta_i - (1-\gamma) + \delta_i x_i^a$. When $\beta_i < 1 - \gamma - \delta_i x_i^a$ we therefore have $x_i^* = 0$. Evaluated at $x_i = \overline{x}$, (6) becomes $-\alpha_i - (1-\gamma) - \delta_i (\overline{x} - x_i^a) < 0$. Finally, as highlighted in the proposition, in any interior solution (6) is equal to 0, and hence

$$\delta_i x_i^* + F(x_i^*) (\alpha_i + \beta_i) = \beta_i - (1-\gamma) + \delta_i x_i^a . \tag{7}$$

■

Proof of corollary 1 Part (i) Suppose that, for some $\widehat{x} \in (0, \overline{x})$, the following holds: $F_1(x) >$

$F_0(x)$ for $\in (0, \widehat{x})$ and $F_1(x) < F_0(x)$ for $x \in (\widehat{x}, \overline{x})$. Consider $x^*_{i0} \in (0, \widehat{x})$. We have

$$\begin{aligned} -(1-\gamma) - \alpha_i F_1\left(x^*_{i0}\right) + \beta_i(1 - F_1\left(x^*_{i0}\right)) - \delta_i\left(x_i - x^a_i\right) &< \\ -(1-\gamma) - \alpha_i F_0\left(x^*_{i0}\right) + \beta_i(1 - F_0\left(x^*_{i0}\right)) - \delta_i\left(x_i - x^a_i\right) &= 0 \end{aligned}$$

Since $-(1-\gamma) - \alpha_i F\left(x\right) + \beta_i(1 - F\left(x\right)) - \delta_i\left(x - x^a_i\right)$ is decreasing in x, this implies that $x^*_{i1} < x^*_{i0}$. The proof of part (ii) is analogous and is therefore omitted. ■

Proof of corollary 2. From (7) we see that, restricting attention to interior solutions, we have

$$\begin{aligned} \frac{dx^*_i}{d\alpha_i} &= -\frac{F(x^*_i)}{\delta_i + (\alpha_i + \beta_i)f(x^*_i)} < 0 \\ \frac{dx^*_i}{d\beta_i} &= \frac{1 - F(x^*_i)}{\delta_i + (\alpha_i + \beta_i)f(x^*_i)} > 0 \\ \frac{dx^*_i}{dx^a_i} &= \frac{\delta_i}{\delta_i + (\alpha_i + \beta_i)f(x^*_i)} > 0 \end{aligned}$$

■

Proof of corollary 3 Consider first the difference in α necessary to induce a change in x^*_i from x' to x''. Fix the other parameters to be equl to β, δ and x^a and denote as α'_k the value of α required for x' to be optimal under f_k. We have

$$F_k\left(x'\right) = \frac{\varkappa}{\alpha'_k + \beta}$$

where $\varkappa \equiv \beta - (1-\gamma) + \delta x^a - \delta x'$, and, analogously,

$$F_k\left(x''\right) = \frac{\varkappa}{\alpha''_k + \beta}.$$

Recall that $F_1\left(x'\right) > F_0\left(x'\right)$ while $F_1\left(x''\right) < F_0\left(x''\right)$. As a result,

$$\begin{aligned} \frac{\varkappa}{\alpha'_0 + \beta} &< \frac{\varkappa}{\alpha'_1 + \beta} \text{ i.e. } \alpha'_1 < \alpha'_0 \\ \text{and similarly } \frac{\varkappa}{\alpha''_0 + \beta} &> \frac{\varkappa}{\alpha''_1 + \beta} \text{ i.e. } \alpha''_1 > \alpha''_0 \end{aligned}$$

which implies that $\alpha'_0 - \alpha''_0 > \alpha'_1 - \alpha''_1$. Consider now the difference in β necessary to induce a change in x^*_i from x' to x''. We have

$$F_k\left(x'\right) = \frac{\beta'_k + \varrho(x')}{\alpha + \beta'_k}$$

where $\varrho(x) \equiv -(1-\gamma) + \delta x^a - \delta x$, and, analogously,

$$F_k\left(x''\right) = \frac{\beta''_k + \varrho}{\alpha + \beta''_k}.$$

Recalling that $F_1(x') > F_0(x')$ while $F_1(x'') < F_0(x'')$,

$$\begin{array}{rccl} & \dfrac{\beta_0' + \varrho(x')}{\alpha + \beta_0'} & < & \dfrac{\beta_1' + \varrho(x')}{\alpha + \beta_1'} \text{ i.e. } \beta_0'\left(\alpha - \varrho(x')\right) < \beta_1'\left(\alpha - \varrho(x')\right) \\ \text{and similarly} & \dfrac{\beta_0'' + \varrho(x'')}{\alpha + \beta_0''} & > & \dfrac{\beta_1'' + \varrho(x'')}{\alpha + \beta_1''} \text{ i.e. } \beta_0''\left(\alpha - \varrho(x'')\right) > \beta_1''\left(\alpha - \varrho(x'')\right). \end{array}$$

where $\alpha - \varrho > 0$. To see this note that, from (7),

$$\alpha_i > \beta_i(1 - F(x_i^*)) - (1-\gamma) + \delta x^a - \delta x_i^* > \varrho(x_i^*).$$

This implies that $\beta_0'' - \beta_0' > \beta_1'' - \beta_1'$. Finally, the result with respect to x_i^a is derived analogously to α and β and is therefore omitted.■

A.1 The disutility of deviating from x_i^a takes the absolute value form

Suppose that the disutility from selecting a contribution that differs from x_i^a is given by

$$\begin{array}{c} -\vartheta_i(x_i - x_i^a) \text{ if } x_i > x_i^a \\ -\rho_i(x_i^a - x_i) \text{ if } x_i < x_i^a \end{array}$$

for some $\vartheta_i \geq 0$ and $\rho_i \geq 0$. If $x_i > x_i^a$, individual expected utility is

$$X - x_i + \gamma(x_i + x_j) - \alpha_i \int_0^{x_i} (x_i - x_j) f(x_j)dx_j - \beta_i \int_{x_i}^{\overline{x}} (x_j - x_i) f(x_j)dx_j - \vartheta_i(x_i - x_i^a)$$

while if $x_i < x_i^a$ it is

$$X - x_i + \gamma(x_i + x_j) - \alpha_i \int_0^{x_i} (x_i - x_j) f(x_j)dx_j - \beta_i \int_{x_i}^{\overline{x}} (x_j - x_i) f(x_j)dx_j - \rho_i(x_i^a - x_i)$$

The optimal contribution x_i^* satisfies

$$\begin{array}{ll} F(x_i^*) = \frac{\beta_i - (1-\gamma) - \vartheta_i}{\alpha_i + \beta_i} & \text{if } \frac{\beta_i - (1-\gamma) - \vartheta_i}{\alpha_i + \beta_i} > F(x_i^a) \\ F(x_i^*) = \frac{\beta_i - (1-\gamma) + \rho_i}{\alpha_i + \beta_i} & \text{if } \frac{\beta_i - (1-\gamma) + \rho_i}{\alpha_i + \beta_i} < F(x_i^a) \\ x_i^* = x_i^a & \text{if } \frac{\beta_i - (1-\gamma) + \rho_i}{\alpha_i + \beta_i} > F(x_i^a) > \frac{\beta_i - (1-\gamma) - \vartheta_i}{\alpha_i + \beta_i} \end{array}$$

Consider two distributions f_0 and f_1 with the same mean and suppose that f_0 is single-crossing stochastic dominant over f_1 so that, for some $\widehat{x} \in (0, \overline{x})$ the following holds: $F_1(x) > F_0(x)$ for $x < \widehat{x}$ and $F_1(x) < F_0(x)$ for $x > \widehat{x}$.

Clearly enough, the results outlined in corollaries 1 and 2 continue to hold. Consider now corollary 3. It is straightforward to check that $\alpha_0' - \alpha_0'' > \alpha_1' - \alpha_1''$ and $\beta_0'' - \beta_0' > \beta_1'' - \beta_1'$ as in the proof of corollary 3. This implies that the the parameter shift in α_i or in β_i needed to generate a change in optimal contribution from x' to x'' is larger under f_0 compared to f_1. However, the prediction with respect to x_i^a is less clear-cut. Suppose for instance that $\frac{\beta_i - (1-\gamma) - \vartheta_i}{\alpha_i + \beta_i} = x'$ and $\frac{\beta_i - (1-\gamma) + \rho_i}{\alpha_i + \beta_i} = x''$. A change in x_i^a from $\widetilde{x}' = F^{-1}(x')$ to $\widetilde{x}'' = F^{-1}(x'')$ would result in the optimal contribution moving from x' to x''. Since $F_1^{-1}(x') < F_0^{-1}(x')$ and $F_1^{-1}(x'') > F_0^{-1}(x'')$,

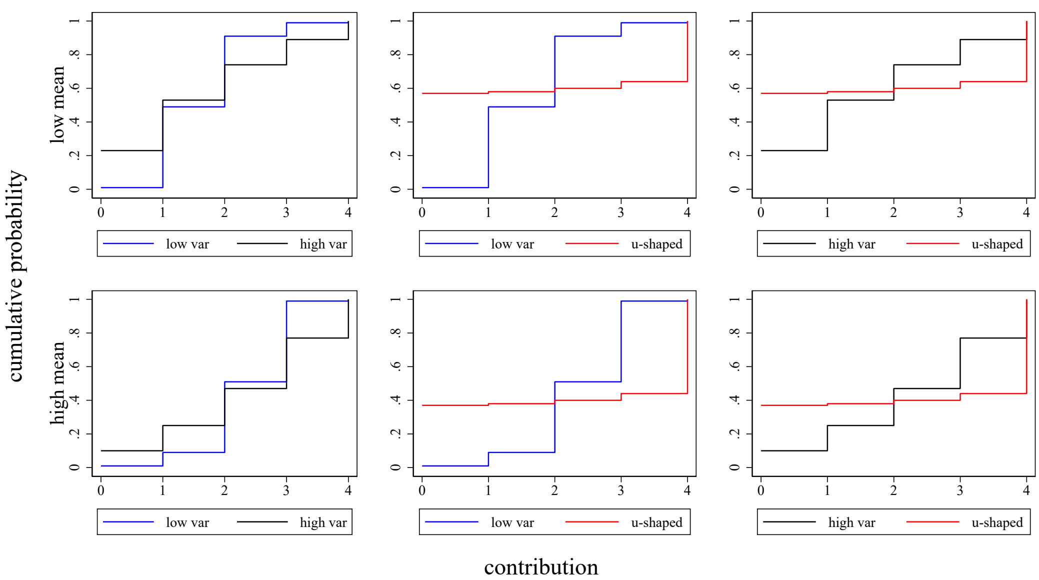
$\widetilde{x}'' - \widetilde{x}'$ is larger under f_1 compared to f_0. We conclude that, if the disutility of deviating from x_i^a takes the absolute value form, the result of corollary 3 does not hold for x_i^a (but it does for α_i and β_i).

B Additional analysis

B.1 Single crossing property of shown distributions

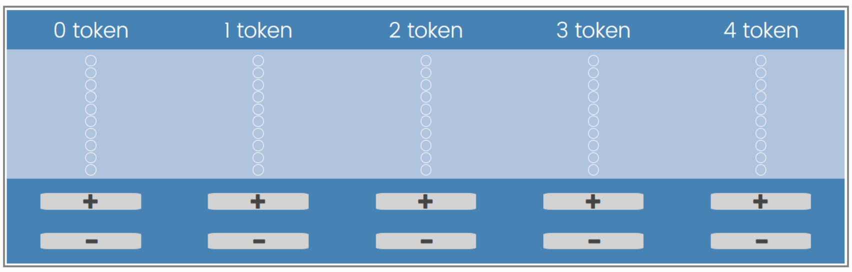
Figure B.1 plots the cumulative distributions of all six treatments. As can be seen below, the single-crossing condition discussed in Section 2 holds for all pairwise comparisons of descriptive norms sharing the same mean.

Figure B.1: Cumulative distributions of descriptive norm treatments



B.2 Belief elicitation screen

Figure B.2: Belief elicitation screen

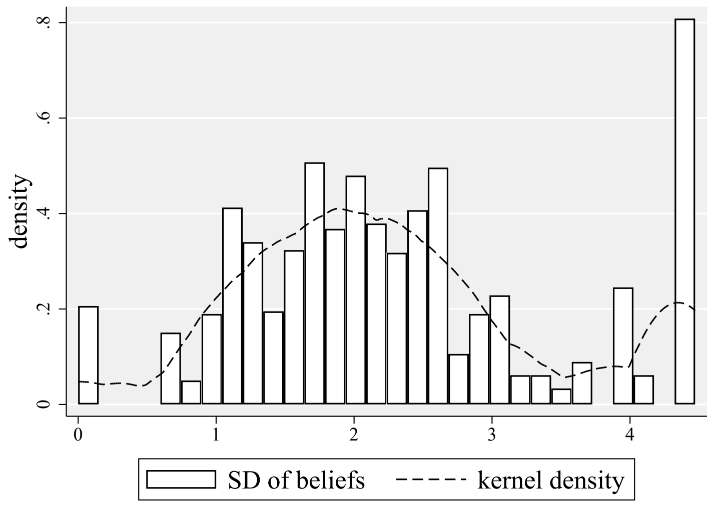


Note. Participants have to allocate a total of 10 points across all available options. The more likely they think an option is the more points they should allocate to it.

B.3 Variation of beliefs

Figure 5 in the main text shows the aggregate distribution of beliefs about the co-player's contribution in Part I, with most people believing that the other will either contribute 2 or 4 tokens. This aggregate distribution hides a substantial heterogeneity between participants in the distribution of beliefs. Figure B.3 shows that the standard deviation of initial beliefs varies substantially between participants (min=0, max=4.5).

Figure B.3: Variation in standard deviations of beliefs in Part I

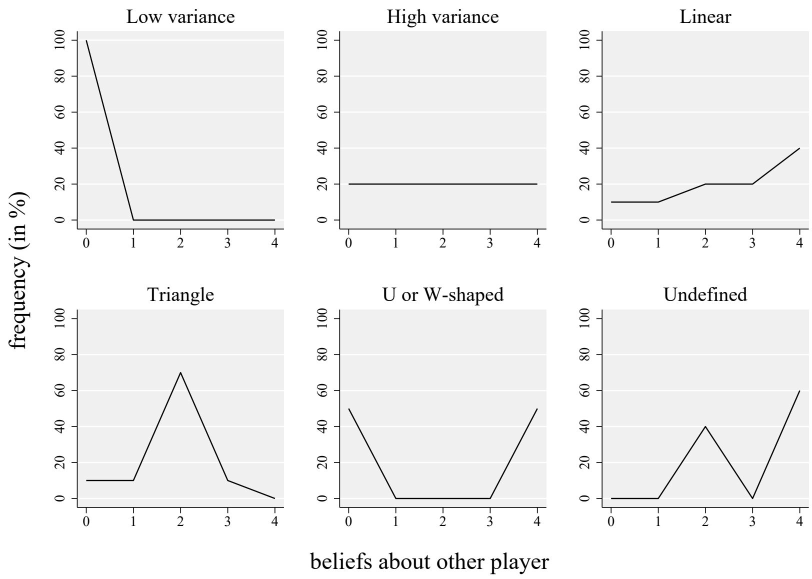


The individual distributions of beliefs can roughly be categorised into 6 types:

1. Low variance: participants who put more than 80% of points on one single option
2. High variance: participants who believe outcomes are equally likely (10 - 30% per option)
3. Linear: participants who have either non-decreasing or non-increasing beliefs
4. Triangle: participants with a single modal belief below 80% that is 1,2, or 3 token
5. U- or W-shaped: participants with either two modes (at 0,4) or three modes (at 0,2,4)
6. Others: not defined by the previous categories

Figure B.4 provides examples for each type. Using these rules we can classify 92% of participants. The most common types are triangles (41%), followed by low variance (21%), u-/w-shaped (15%), linear (10%), and high variance types (5%).

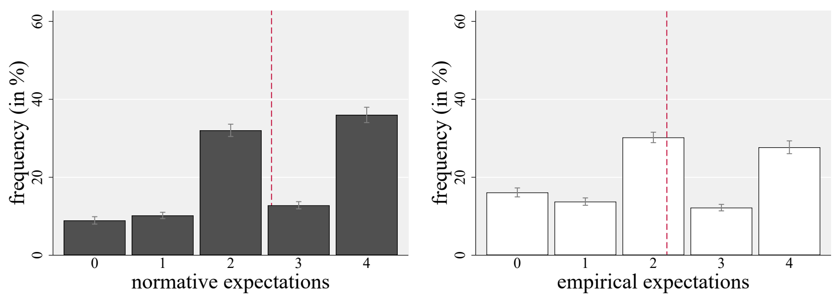
Figure B.4: Examples of individual distributions of beliefs in Part I



B.4 Empirical and normative expectations

Empirical (EE) and normative (NE) expectations in Part I are very similar to contributions, beliefs and PVs (see Figure B.5). Not surprisingly, when asked about what most other people actually contribute (EEs) participants answer in the same way as when asked about their co-player's likely contribution. NEs, are shifted slightly to the right, indicating that people think others contribute less than they say one should. In addition, to eliciting the distribution of expectations and beliefs, we measure the participants' confidence in their replies. Interestingly, participants also appear to be more certain about what others say one should contribute than actual contributions. Although the difference is small, our confidence measure is significantly higher for NEs than EEs about the other player (Wilcoxon signed-rank test, $p = 0.01$).

Figure B.5: NEs and EEs in Part I

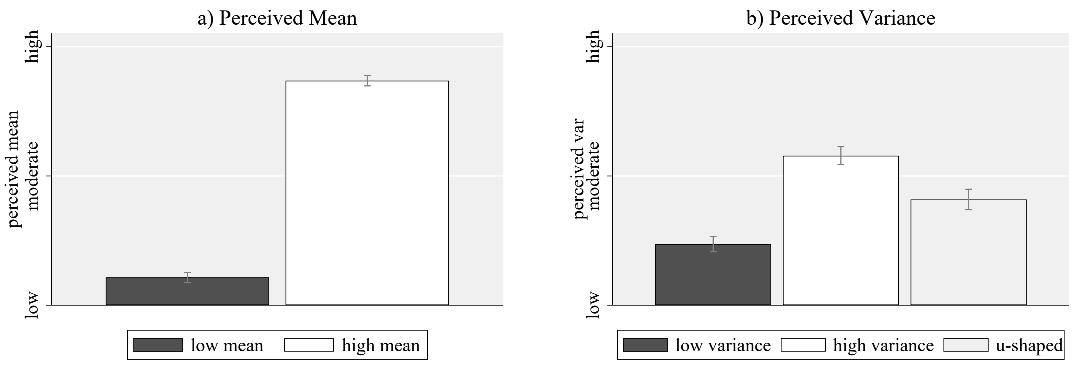


Note. The dashed lines represent averages. Whiskers show 95% confidence intervals.

B.5 Manipulation checks

In the ex-post survey we ask participants about their perceptions of the shown distribution. As Figure B.6a shows, participants in high mean conditions also have a significantly higher perception of the mean (Wilcoxon-Mann-Whitney test, $p < 0.001$). In terms of the variance (Figure B.6b), participants perceive the low variance condition as significantly less varying than the high variance and the u-shaped conditions, and the u-shaped as less varying than the high variance condition (Wilcoxon-Mann-Whitney tests, for all $p < 0.001$). In addition, we ask participants about their difficulty in interpreting the distribution and how common they think this distribution is in the general population. On a scale from 1 (very easy) to 7 (very difficult) the average difficulty rating is 2.1, indicating that participants do not seem to have trouble interpreting the information. Moreover, there is no difference in the difficulty rating between high and low mean conditions or u-shaped and low variance conditions. Only the high variance condition is described as significantly harder to understand than both the u-shaped (Wilcoxon-Mann-Whitney tests, $p = 0.03$) or the low mean condition (Wilcoxon-Mann-Whitney tests, $p = 0.08$). However, the difference is very small (0.2 on a scale from 1 to 7). The average rating of how common the shown distributions are perceived to be is 5.0 on a scale from 1 (very rare) to 7 (very common). Again, there are no differences between high and low mean conditions or u-shaped and high/low variance conditions. Only the high variance condition is perceived as slightly less common than the low variance condition (Wilcoxon-Mann-Whitney tests, $p = 0.009$). Again, the difference is extremely small (0.2).

Figure B.6: Manipulation checks



Note. Whiskers show 95% confidence intervals.

B.6 Effect of variance and shape of the distribution

Figure B.7 shows the distribution of contributions in Part II by variance/shape, but pooled over low and high mean conditions. As described in the main text, we can see that the variance of contributions is largest for the u-shaped conditions, followed by the loose/ high variance conditions, while there is less variance in the tight/ low variance conditions.

As our theoretical framework assigns a particular role to reciprocity concerns and conditionality, analysing the effect of variance is particularly relevant for conditional contributors. As 84% of our sample are conditional cooperators, we do not analyse types separately in the

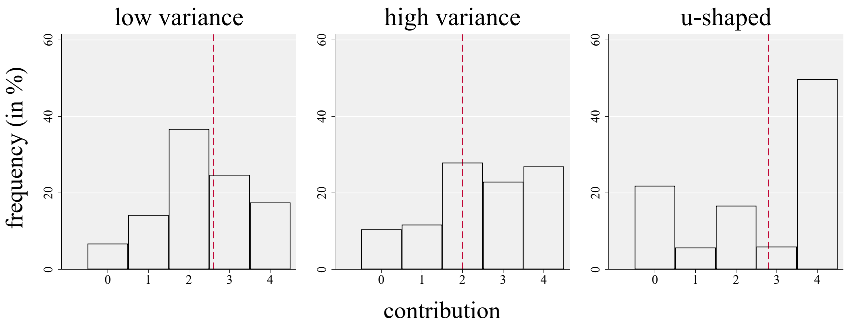


Figure B.7: Distribution of participants' contribution in Part II by variance

main paper and report results for all participants. Figure B.8 shows that when only focusing on conditional cooperators, our results hold: tight, loose, and polarized environments induce tight, loose, and polarized responses. We also confirm that the variance of contributions in Part II among conditional cooperators in the u-shaped condition is significantly larger than in the high or low variance condition (F-tests, $p < 0.001$). The variance of contributions in the high variance conditions in turn is significantly larger than in the low variance conditions (F-test, $p < 0.001$). Also the distributions are significantly different from each other.

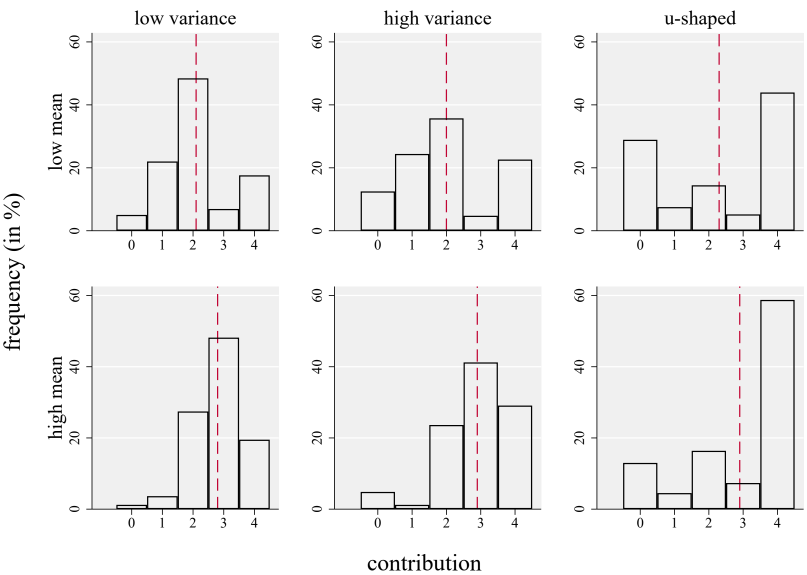
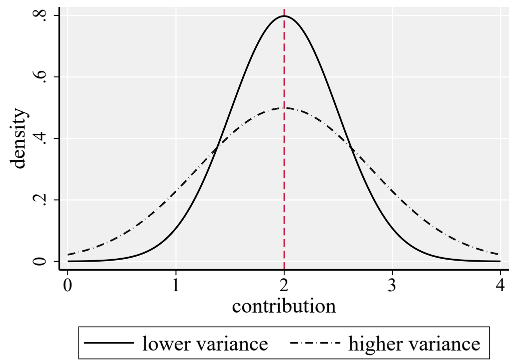


Figure B.8: Distribution of participants' contribution in Part II (conditional cooperators only)

A more indirect test of whether the variation in contribution increases with the observed variance is by splitting our sample in Part II in a high and a low contribution group (contributions above/ below the median) and comparing their reactions to different environments. If a higher observed variation increases the overall variance in contributions, we would expect this to translate into lower contributions for the *low contribution sub-sample* and higher contributions for the *high contribution sub-sample*, resulting in an overall wider spread of contributions (see Figure B.9).

Table B.1 shows the results of regressing contributions on treatment conditions for each sub-sample. We can see that in fact in the high contribution sub-sample, both the high variance and

Figure B.9: Different variances and their effect on the high/low end of the distribution



Note. As the figure shows a higher variance implies more extreme contributions in both tails of the distribution

the u-shaped conditions lead to a significant increase in contributions relative to the low variance condition. For the low contribution sub-sample, by contrast, we see the exact opposite pattern. In this case, a higher variance is associated with lower contributions. We thus confirm that both the u-shaped and the high variance condition increase the overall variation of contributions.

Table B.1: Tobit models. Effect of variance on contributions in Part II for different sub-samples

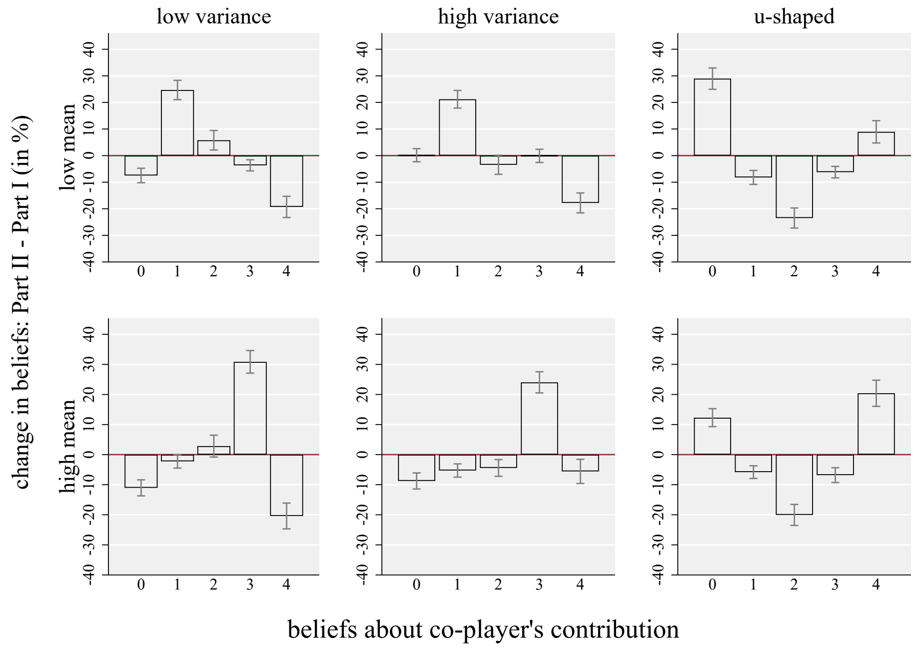
	(1) High contributions (>2)	(2) Low contributions (<=2)
Variance (baseline = low)		
High variance	0.33***	-0.14
	(0.09)	(0.09)
U-shaped	1.20***	-0.77***
	(0.10)	(0.10)
Constant	0.96**	0.27
	(0.44)	(0.42)
Baseline controls	Yes	Yes
Demographic controls	Yes	Yes
N observations	906	602
Pseudo R^2	0.19	0.13

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.
Note. Data is censored at 0 and 4. *Variance* is a categorical variable with 0 = low variance, 1 = high variance and 2 = u-shaped. Baseline controls include contributions, average beliefs, and PVs in Part I and can take values between 0 and 4. Demographic controls include age, gender, education, acceptance of risk, trust sucker aversion, free-riding aversion and measures for negative and positive reciprocity. High and low contribution sub-samples are generated by dividing participants in a group with contributions in Part II above and below the median (=2).

B.7 Changes in beliefs

Figure B.10 shows that participants change their beliefs in line with the shown distributions. For instance, for the u-shaped conditions, the beliefs that the other player contributed 0 or 4 tokens increase substantially between Part I and II of the experiment, while intermediate values (1-3) decrease. The opposite is true for the low and high variance conditions, where in line with the observed distribution, the beliefs that the other contributes 0 or 4 tokens decrease in favor of intermediate values. Intuitively, the additional information in Part II also explains that participants state a significantly higher confidence in their beliefs as compared to Part I (Wilcoxon signed-rank test, $p < 0.001$).

Figure B.10: Changes in beliefs about the co-player's distribution between parts



Note. Whiskers show 95% confidence intervals.

B.8 Personal traits and demographics

Table B.2 shows that the analysis of personal traits from section 4.3 holds when restricting the sample to participants classified as conditional cooperators. We still see that sucker-aversion lowers contributions, while free-riding aversion and personal values have a positive effect on contributions. Most importantly, we confirm that the effect of these personal traits is larger in conditions with higher variance (loose/polarized).

Table B.2: Tobit models. Effect of personal values, sucker, and free-riding aversion on contributions in Part II (conditional cooperators only)

	(1)	(2)
Sucker aversion	-0.20***	-0.19***
	(0.06)	(0.06)
Sucker aversion x high variance	0.03	0.06
	(0.09)	(0.09)
Sucker aversion x u-shaped	-0.22**	-0.21**
	(0.09)	(0.09)
Free-riding aversion	0.12**	0.12**
	(0.05)	(0.05)
Free-riding aversion x high variance	0.08	0.07
	(0.08)	(0.08)
Free-riding aversion x u-shaped	0.28***	0.29***
	(0.08)	(0.08)
Personal values (PVs)	0.43***	0.43***
	(0.10)	(0.10)
PVs x high variance	0.25*	0.25*
	(0.14)	(0.14)
PVs x u-shaped	0.39***	0.40***
	(0.14)	(0.14)
Constant	1.35***	-0.37
	(0.44)	(0.76)
Demographic controls	No	Yes
N observations	1006	994
Pseudo R^2	0.09	0.11

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

Note. Regressions control for mean and variance of the observed distribution. *Personal values* can take values between 0 and 4. Sucker and free-riding aversion are measured on a Likert scale from 1 to 7. Demographic controls include age, gender, education, acceptance of risk, trust, and measures for negative and positive reciprocity. All regressions control for order effects.

We measure personal values both in Part I before participants receive any information about co-player behavior and in Part II after they see the distributions. In the main text, we use personal values in Part I as a regressor as they cannot be affected by the treatments and thus represent a cleaner representation of what participants think is the right thing to do. However, Table B.3 shows that our results hold after controlling for the change in personal values between parts and their interaction with treatment indicators. Note that changes in personal values are very minor. On average participants change their personal values by -0.1 (with personal values taking values between 0 and 4). Although this is statistically different from zero (Wilcoxon signed-rank test, $p < 0.001$), the median change is 0, with 69% of participants not changing their reported personal values between parts.

Table B.3: Tobit models. Effect of personal values, sucker, and free-riding aversion on contributions in Part II (controlling for change in PVs)

	(1)	(2)
Sucker aversion	-0.13**	-0.12**
	(0.05)	(0.05)
Sucker aversion x high variance	-0.06	-0.02
	(0.08)	(0.07)
Sucker aversion x u-shaped	-0.24***	-0.22***
	(0.08)	(0.08)
Free-riding aversion	0.13***	0.14***
	(0.05)	(0.05)
Free-rider aversion x high variance	0.05	0.04
	(0.07)	(0.07)
Free-rider aversion x u-shaped	0.28***	0.28***
	(0.07)	(0.07)
Personal values in part I	0.75***	0.71***
	(0.10)	(0.10)
High variance x PVs in part I	0.23*	0.26**
	(0.13)	(0.13)
PVs in part I x u-shaped	0.41***	0.44***
	(0.14)	(0.14)
Change in personal values	0.56***	0.51***
	(0.11)	(0.11)
Change in PVs x high variance	0.22	0.24
	(0.17)	(0.16)
Change in PVs x u-shaped	0.27*	0.31**
	(0.14)	(0.14)
Constant	0.23	-1.55**
	(0.37)	(0.66)
Demographic controls	No	Yes
N observations	1203	1188
Pseudo R^2	0.14	0.15

* p < 0.1, ** p < 0.05, *** p < 0.01. Standard errors in parentheses.

Note. Regressions control for mean and variance of the observed distribution. *Personal values* are measured in Part I and can take values between 0 and 4. Sucker and free-riding aversion are measured on a Likert scale from 1 to 7. Demographic controls include age, gender, education, acceptance of risk, trust, and measures for negative and positive reciprocity. All regressions control for order effects.

Finally, we find that some demographic variables correlate with personal traits (see Table B.4) with women and younger participants having both a higher sucker and a higher free-riding aversion. Personal values on the other hand seem to be slightly lower for female participants.

Table B.4: Tobit models. Correlates of personal values, sucker and free-riding aversion

	Personal values	Sucker aversion	Free-riding aversion
	(1)	(2)	(3)
Female	-0.14**	0.59***	1.15***
	(0.07)	(0.15)	(0.19)
Age	-0.00	-0.03***	-0.02***
	(0.00)	(0.00)	(0.01)
Education (baseline=no formal degree)			
Secondary school	-0.08	-0.63	-0.62
	(0.23)	(0.54)	(0.67)
University/ college	-0.04	-0.71	-1.13*
	(0.23)	(0.53)	(0.66)
Prefer not to say	-0.68	-0.45	-1.67
	(0.41)	(0.93)	(1.18)
Constant	3.02***	6.52***	5.13***
	(0.24)	(0.56)	(0.69)
N observations	1188	1188	1188
Pseudo R^2	0.002	0.015	0.011

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.
Note. Sucker and free-riding aversion are measured on a Likert scale from 1 to 7. PVs can take values between 0 and 4.

C Quadratic scoring rule

The quadratic scoring rule (QSR) is used to elicit participants' beliefs about uncertain events in an incentive compatible way. As a consequence participants should reveal their beliefs truthfully. The key idea of the QSR is that participants are rewarded for the correct assessment, but penalized for every wrong assessment by the squared distance between their guess and the true event, using the following equation (Murphy and Winkler, 1970):

$$Q_j(p) = \alpha + 2\beta p_j - \beta \sum_{i=1}^{n} (p_i)^2, \tag{8}$$

where p_i is the probability an individual assigns to even i and p_j is the probability an individual assigns to the true event. Artinger et al. (2010) develop and test a salient and intuitive way to communicate information about payoffs to participants using a decomposed version of the equation above:

$$Q_j(p) = (\alpha + \beta) - \beta(1 - p_j)^2 - \beta \sum_{i \neq j} (p_i)^2, \tag{9}$$

The first two terms represent the earnings from the correct option. Participants are penalized for not assigning a probability of 1 to the correct event $(-\beta(1-p_j)^2)$. In the experiment we set $\alpha = \beta = \$0.5$. This implies that if an individual assigns full probability to the correct event they earn \$1, if they assign a zero probability to the correct event they earn \$0.5 from the correct option. The last term in the equation above represents the penalty for assigning a positive probability to events that do not occur. If individuals assign zero probability to untrue events, this penalty is zero. If they assign all probabilities to wrong events, the penalty is \$0.50. This means that the maximum amount individuals can earn in the elicitation task is \$1, while the minimum is \$0.

Artinger et al. (2010) show that if payoffs are split up in this way, this significantly facilitates participants' understanding of the QSR and the implications for payoffs. Figure C.1 shows how the QSR is presented to participants.

Figure C.1: Representation of QSR

Points put on the correct option	**Earnings from the correct option**	**Points put on a wrong option**	**Costs from a wrong option**
10	100¢	10	50¢
9	99.5¢	9	40.5¢
8	98¢	8	32¢
7	95.5¢	7	24.5¢
6	92¢	6	18¢
5	87.5¢	5	12.5¢
4	82¢	4	8¢
3	75.5¢	3	4.5¢
2	68¢	2	2¢
1	59.5¢	1	0.5¢
0	50¢	0	0¢

D Instructions

Welcome

Thank you very much for participating in this study! This study consists of two parts and a questionnaire. Upon completion you will receive \$1.70 for your participation plus an additional bonus of up to \$2.36 that depends both on your decisions and the decisions of other participants. **In both parts you will face a situation in which you will be matched with one other, real participant**. On the next page we will describe this situation to you in more detail.

Instructions (1/2)

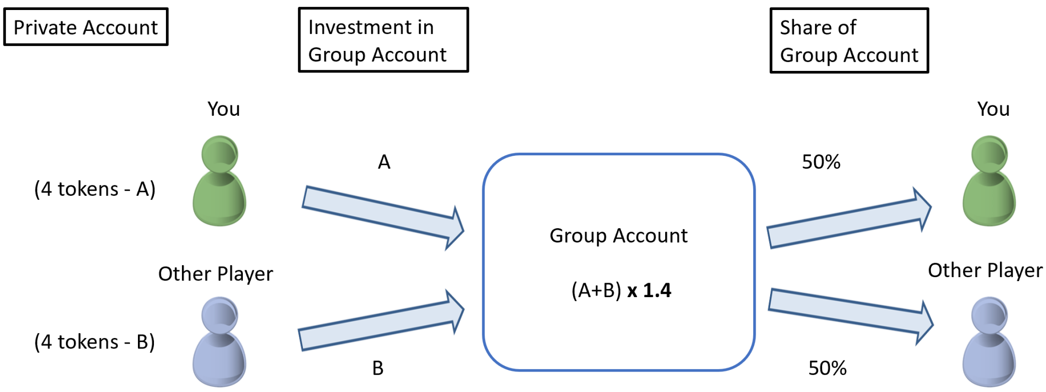
In this study, you will be anonymously paired with another participant. You will each start with **4 token** in your personal **private accounts**. In addition to the private accounts, there is a **group account**. You have to decide how many of your token you want to invest in the group account (either 0, 1, 2, 3 or 4 token). The amount leftover will remain in your private account. The other player has to make the same decision.

Your income from the private account

The amount in the private account is yours to keep. The other player doesn't earn anything from the token you keep in your private account. For example, if you keep 2 token in your private account, this will be your income from this account.

Your income from the group account

The amount invested in the group account will be multiplied by 1.4. That is, each token invested in the group account will yield 1.4 token for the group. The total amount in the group account will be split equally between you and your partner regardless of your individual investments. That is, each player receives half (50%) of the total amount in the group account.



If, for example, the sum of all investments in the group account by you and the other player is 6 token (A+B), then the group account yields 6 x 1.4 = 8.4 token. Both you and the other player would then receive 0.5 x 8.4 = 4.2 token from this account.

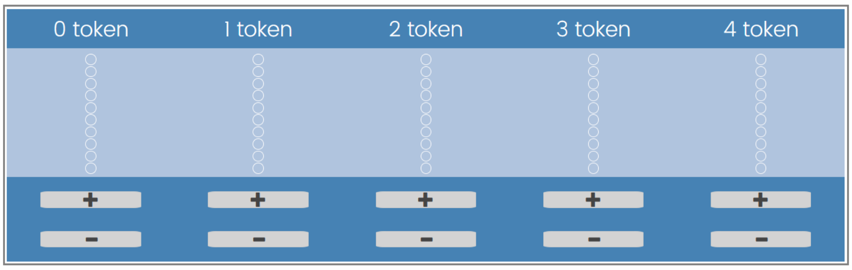
Your total income

Please note that for logistic reasons you are not interacting with other participants in real time. Once we collected all responses, we will match you with another person to calculate your and

the other's total income.[27] The latter consists of **all the token you kept in your private account plus half of the token** that you and that other participant invested **in the group account**.
Your total income will then determine your bonus payment, with each token being worth \$0.10. Whatever income you earned in token will be converted at this rate into actual money at the end of the experiment and paid out as a bonus.

Instructions (2/2)

In addition to making an investment decision in this situation, we will ask you to state **your beliefs about other participants**. You will be paid for these tasks according to how accurate your beliefs are.
A brief explanation follows: let us assume, we ask you to make a guess about how many token the participant you have been matched with invested in the group account. In this case, you would have to indicate **how likely you think it is that the other participant invested 0, 1, 2, 3 or 4 token**. To make your choices you will see a screen like the one below.



To make your decision you have to allocate a total of 10 points across options by clicking on the plus and minus buttons. The points you allocate need to add up to 10 and **the more likely you think one option is, the more points you would allocate to it**. The points you allocate to each option will naturally reflect your beliefs about the other participant's behavior.
The amount of money you can earn depends on how you allocated your points and what is actually true. If you put all points on the correct option, you will earn \$1 if you put all points on a wrong option you will earn \$0. In general, the more points you allocate to a correct option, the higher your earnings and the more points you allocate to a wrong option the lower your earnings. **The way your earnings are determined ensures that your best strategy is to carefully and honestly answer these questions.** If you want to have a closer look at how your earnings will be calculated click **here**.
Let's for example assume that you think it is equally likely that the other participant invested 2 or 4 token and you put 5 points on each option. If the other participant really invested either 2 or 4 token, you would in each case earn \$0.75. If they invested 0, 1 or 3 token you would earn \$0.

[27] Note that this is a common set-up for studies on Prolific and thus familiar to participants. The Prolific guidelines allow up to 21 days to pay participants.

What if you had instead put all your eggs in one basket and allocated 10 points on the other participant investing 2 token? If the other participant indeed invested 2 token, you earn the maximum bonus of $1. But if any of the other options is the correct one, you would earn nothing in this task. It is thus up to you to balance the strength of your personal beliefs with the risk of them being wrong.

In total, we will ask you to state your belief on **five** different questions throughout this study. In the end, a lottery will decide **one of them to be chosen for payment**. The amount you earned in the chosen question will then be added to your bonus payment.

If participants click to get more information about payoffs, they see the following pop-up:

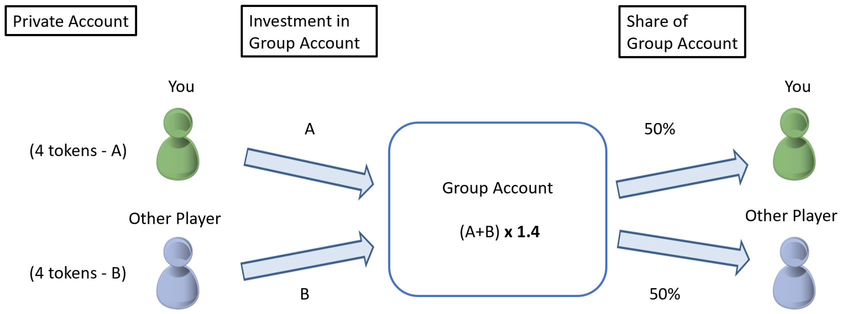
Your earnings are calculated on the basis of the table below. The more points you put on the correct option the higher your earnings. For each wrong option to which you allocate points your earnings will be reduced. The reduction is larger the more points you allocated to that option.

Points put on the correct option	**Earnings from the correct option**	**Points put on a wrong option**	**Costs from a wrong option**
10	100¢	10	50¢
9	99.5¢	9	40.5¢
8	98¢	8	32¢
7	95.5¢	7	24.5¢
6	92¢	6	18¢
5	87.5¢	5	12.5¢
4	82¢	4	8¢
3	75.5¢	3	4.5¢
2	68¢	2	2¢
1	59.5¢	1	0.5¢
0	50¢	0	0¢

Part 1

We are now going to ask you a number of questions that relate to the situation that you previously read (see image below). It is important that you answer these questions truthfully and as accurately as possible.[28]

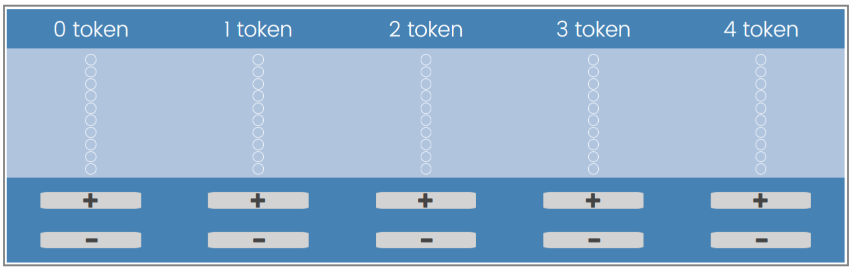
[28] Either normative questions or ABC questions are asked first. The three normative questions appear in randomized order.



1) We asked other participants what they believe is the **most appropriate amount to invest** in the **group account**. What do you believe was the most common answer?

Appropriate here means what you personally consider to be "correct" or "moral". You can allocate up to 10 points to an option by clicking on the plus and minus buttons below to indicate your guess. The more likely you think one option is, the more points you would allocate to it. Your guesses need to add up to 10.

Most people believe it is appropriate to invest...

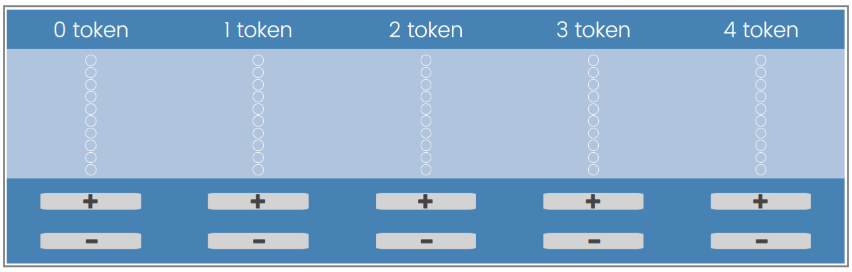


How confident are you in your response above? (0 not very confident, 100 very confident)

2) We asked other participants to make an investment decision in this situation. How many token do you believe most people **actually invested** in the group account?

You can allocate up to 10 points to an option by clicking on the plus and minus buttons below to indicate your guess. The more likely you think one option is, the more points you would allocate to it. Your guesses need to add up to 10.

Most people actually invested...



How confident are you in your response above? (0 not very confident, 100 very confident)

3) According to your own opinion and independent of the opinion of others, what is the **most appropriate amount to invest** in the **group account**?

Appropriate here means what you personally consider to be "correct" or "moral".

- 0 token
- 1 token
- 2 token
- 3 token
- 4 token

For the next section of Part 1, you will be matched with one other participant. You will only interact once with this person and you will never learn each other's identity.
Your and the other participant's bonus payment for Part 1 will depend on your decisions and the decisions of this participant.

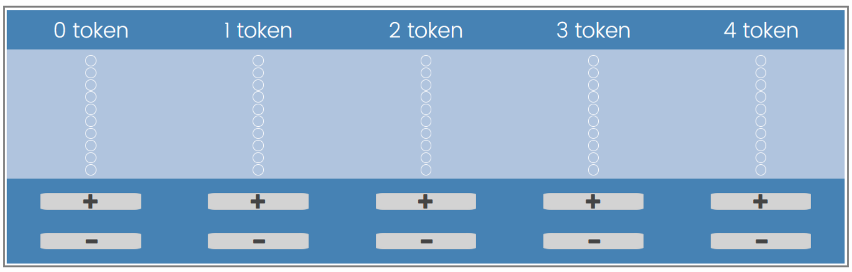
1) How many token do you want to **invest** in the **group account**?

- 0 token
- 1 token
- 2 token
- 3 token
- 4 token

2) How many token do you believe the participant you are matched with **invested** in the **group account**?

You can allocate up to 10 points to an option by clicking on the plus and minus buttons below to indicate your guess. The more likely you think one option is, the more points you would allocate to it. Your guesses need to add up to 10.

The participant you are matched with invested...



How confident are you in your response above? (0 not very confident, 100 very confident)

3) We are also interested in how many token you want to invest in the group account if you could know the other's choice beforehand. This means **you can condition** your investment on your group member's choice.

For one of you, the **unconditional choice** that you took before will count as the investment decision. For the other, the **conditional choice** (according to the table below) will count as the investment decision. Should the conditional choice be selected for you and the other participant invested x token in their unconditional choice, your decision for that scenario will determine your investment and thus matter for your bonus.

To determine your conditional choice, please tell us what you **want to invest** in the **group account** if:

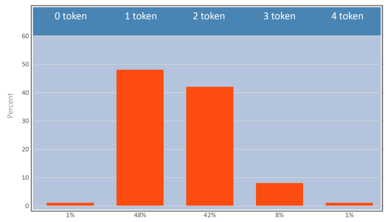
- The other player invests 0 token: ______ token
- The other player invests 1 token: ______ token
- The other player invests 2 token: ______ token
- The other player invests 3 token: ______ token
- The other player invests 4 token : ______ token

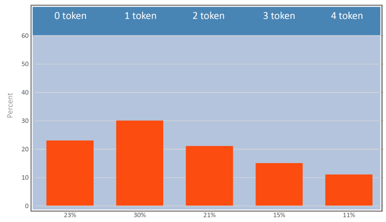
Part 2

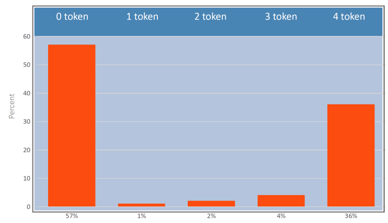
In a previous study we asked over 600 participants to **make an investment decision** in the same situation. The possible choices were to invest 0, 1, 2, 3 or 4 token in the **group account**. From their answers we constructed different sub-groups. The graph below shows the percentage of people choosing each option in one randomly selected sub-group.

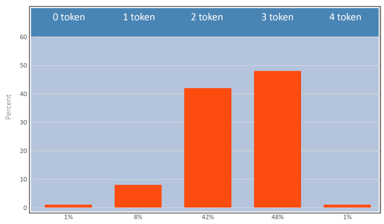
What previous participants invested in the group account:

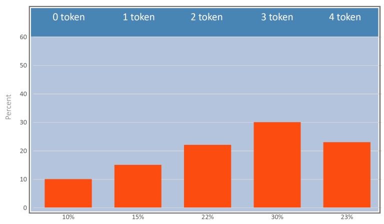
(Participants are randomly shown one of the following six pictures.)

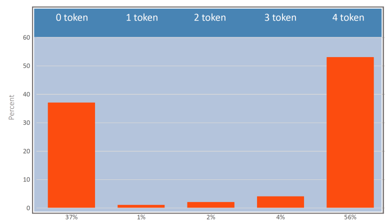












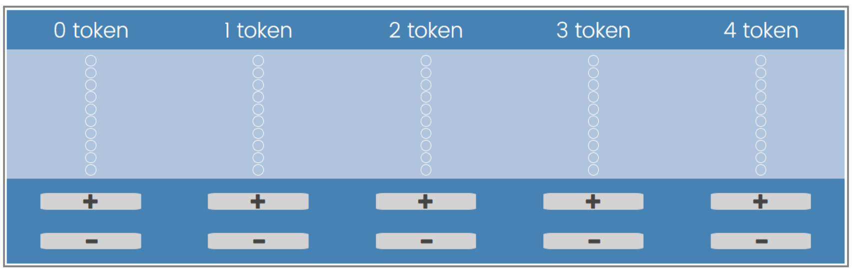
For Part 2 of the experiment you are matched with **one of the participants from the sub-group above**. You will only interact once with this person and you will never learn each other's identity.

Your and the other participant's **bonus payment for Part 2** will depend on your decisions and the decisions of this participant.[29]

1) We asked other participants from the previous study what they believe is the **most appropriate amount to invest** in the **group account**. What do you believe was the most common answer?

Appropriate here means what you personally consider to be "correct" or "moral". You can allocate up to 10 points to an option by clicking on the plus and minus buttons below to indicate your guess. The more likely you think one option is, the more points you would allocate to it. Your guesses need to add up to 10.

Most people believe it is appropriate to invest...



[29] We randomize whether participants are first asked about contributions and beliefs or about personal values and normative expectations.

How confident are you in your response above? (0 not very confident, 100 very confident)

2) Think again about the investment decision itself. According to your own opinion and independent of the opinion of others, what is the **most appropriate amount to invest** in the **group account**?

Appropriate here means what you personally consider to be "correct" or "moral".

- 0 token
- 1 token
- 2 token
- 3 token
- 4 token

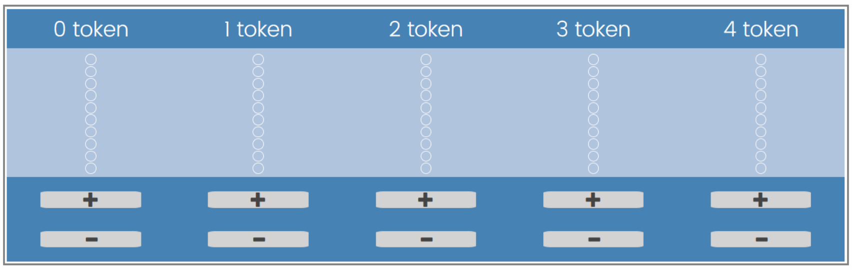
3) How many token do you want to **invest** in the **group account**?

- 0 token
- 1 token
- 2 token
- 3 token
- 4 token

4) How many token do you believe the participant you are matched with **invested** in the **group account**?

You can allocate up to 10 points to an option by clicking on the plus and minus buttons below to indicate your guess. The more likely you think one option is, the more points you would allocate to it. Your guesses need to add up to 10.

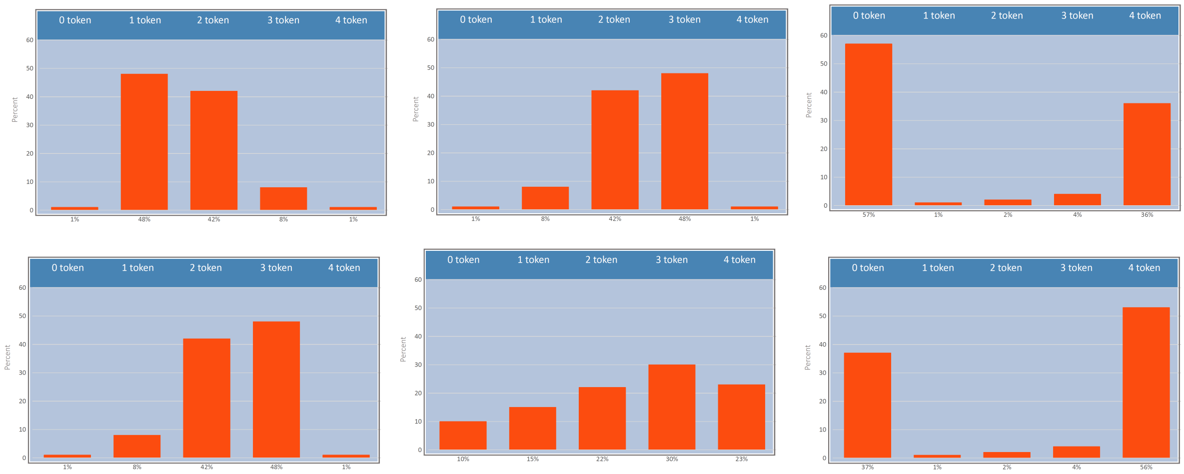
The participant you are matched with invested...



How confident are you in your response above? (0 not very confident, 100 very confident)

Questionnaire (1/2)

In this survey we showed you how many token a sub-group of other participants **invested** in the **group account**. Their answers are represented by the graph below.
What previous participants invested in the group account: *(Participants are randomly shown one of the following six pictures.)*



We will now ask you a few questions about the graph.

1) What are your thoughts on the behavior shown above?

---

2) Would you say the graph shows that overall

- people invest most of their token in the group account
- people invest half of their token in the group account
- people keep most of their token in their private account

3) Would you say the graph shows that overall

- there is a strong tendency for people to invest similar amounts in the group account
- there is a moderate tendency for people to invest similar amounts in the group account
- investments in the group account are very mixed

4) How common do you think the distribution of behavior shown above would be in other groups? (1 very rare, 7 very common)

5) How difficult was it for you to interpret the graph in Part 2, which is also shown above? (1 very easy, 7 very difficult)

6) How upset would you be if you invested everything in the group account and discovered that the participant you have been matched with invested nothing? (1 not at all upset, 7 very upset)

7) How ashamed would you be if you invested nothing in the group account and discovered that the participant you have been matched with invested everything? (1 not at all ashamed, 7 very ashamed)

Questionnaire (2/2)

1) Generally speaking, would you say that most people can be trusted or that you need to be very careful in dealing with people?

- Need to be very careful
- Don't know
- Most people can be trusted

2) In how far do you agree with the following statement: "When someone does me a favour, I will return it." (1 don't agree at all, 7 completely agree)

3) In how far do you agree with the following statement: "If I am treated very unjustly, I will take revenge, even if there is a cost to do so." (1 don't agree at all, 7 completely agree)

4) Please tell me, in general, how willing or unwilling you are to take risks. (1 very unwilling to take risks, 11 very willing to take risks)

5) What is your age? ______ token

6) Which gender do you identify with?

- Female
- Male
- Non-binary
- Other
- Prefer not to say

7) What is the highest level of schooling you completed?

- No formal qualifications
- Secondary school

- University/ college degree
- Prefer not to say

Thanks a lot for participating in this survey! If you have any feedback for us you can write it here: