Strategic Behavior with Tight, Loose and Polarized Norms

Eugen Dimant[†] Michele Gelfand[‡] Anna Hochleitner[§] Silvia Sonderegger[¶]

November 2, 2023

Abstract

Descriptive norms – the behavior of other individuals in one's reference group – play a key role in shaping individual decisions in managerial contexts and beyond. Organizations are increasingly using information about descriptive norms to nudge positive behavior change. When characterizing peer decisions, 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. Furthermore, in polarized and loose environments, personal traits and values play a larger role in determining actual behavior. These nuances of how individuals react to different types of descriptive norms have important implications for company culture, productivity, and organizational effectiveness alike.

Keywords: Cooperation, Descriptive Norms, Variance, Peer Effects
JEL Codes: C91, D01

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, the SITE 2023 conference at Stanford, the 2023 Advances with Field Experiments (AFE) Conference at the University of Chicago, as well as seminars in Cologne, Essex, Konstanz, Montpellier, Hamburg, Leicester, Nottingham, Oxford, and the WZB Berlin. We also thank Zvonimir Bašić, 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.

[†]University of Pennsylvania, USA. CESifo, Germany. Email: edimant@sas.upenn.edu.
[‡]Stanford University, USA. Email: gelfand1@stanford.edu.
[§]NHH (FAIR and SNF), Norway. Email: anna.hochleitner@snf.no.
[¶]University of Nottingham, U.K. Email: silvia.sonderegger@nottingham.ac.uk.

1 Introduction

Descriptive norms - the behavior of other individuals in one's reference group - play a key role in shaping individual decisions across a variety of contexts, from everyday ephemeral interactions to workplace environments and managerial decisions (see for instance Bicchieri, 2005; Sliwka, 2007; Chen et al., 2010; Kimbrough and Vostroknutov, 2016; Danilov and Sliwka, 2017; Fehr and Schurtenberger, 2018; Bicchieri et al., 2022b; Bicchieri and Dimant, 2023). A flourishing line of research in the organizational sciences has shown that descriptive norms affect many aspects of organizational life, including workplace safety, absenteeism, civility, health and wellness, sustainability, entrepreneurship, and ethical conduct, among other domains (Xie and Johns, 2000; Schultz et al., 2008; Stephan and Uhlaner, 2010; Robertson and Barling, 2013; Pek et al., 2017; Jacobson et al., 2020; Belle and Cantarelli, 2021; Taylor et al., 2022). Businesses and organizations can also actively capitalize on social norms for a variety of purposes, including improved customer interaction as well as incorporating employee performance incentives (Huck et al., 2012; Burtch et al., 2018; Ai et al., 2023; Cullen and Perez-Truglia, 2023). This broad evidence suggests that studying how individuals react to the *specific* features of surrounding descriptive norms is essential from a managerial perspective due to their significant impact on employee behavior and organizational outcomes (Burks and Krupka, 2012; Guiso et al., 2015). For instance, understanding how a merger may affect the culture/descriptive norm of the merging organizations is considered crucial for predicting merger success (Weber and Camerer, 2003).

Like many areas of science, there exists a dominant paradigm through which descriptive norms are studied in organizations. When characterizing how individuals behave, a common approach in the existing literature is to focus on *mean* or *modal* behavior (see e.g. Kandel and Lazear, 1992; Grout et al., 2015; Feldhaus et al., 2019).[1] In this paper, we argue that such a focus fails to account for features – the *variance* or *shape* of the distribution of behavior – that play an important role in how people react to a descriptive norm in the presence of strategic interactions. To see this, consider for instance a collective action problem and suppose that individuals in a given community contribute an average of 2 out 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 is inserted into one of these communities and has to interact with a random partner. 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 community. How much should the agent contribute? A mean-focused approach suggests that their contribution should be the same, independently of which descriptive norm the agent is confronted with. Crucially, although these scenarios generate the same average contribution, they clearly depict different social environments, which may generate distinct reactions.

To study this, we vary the *mean* and *variance* of the descriptive norm as well as another

[1] 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.

important feature that characterizes the distribution of behavior, namely its *shape* (unimodal vs. u-shaped vs. polarized). Understanding how people react to polarized environments has become increasingly important, especially in the face of rising political polarization and its detrimental societal outcomes (Iyengar and Westwood, 2015; Iyengar et al., 2019; Enke, 2020; Bauer et al., 2022; Callander and Carbajal, 2022; Dimant et al., 2022; Robbett and Matthews, 2022; Nunn, 2022; Dimant, 2023b; Ren et al., 2023). The role of variance in characterizing descriptive norms is emphasized in the seminal work 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. 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. Variations in the tightness of norms have also been linked with cross-cultural differences in norm-following and norm stability (Aycinena et al., 2022; Kimbrough et al., 2022).

With that, 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 that is ubiquitous in all environments involving social interactions: public goods provision. 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; Fischbacher and Gächter, 2010; Bowles and Gintis, 2013; Kölle and Quercia, 2021).[2] 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, in our experiment as well as in real life people 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 polarized or loose environments with high behavioral variance, strategic uncertainty is high since it is hard to predict how any given individual will behave. In very tight environments, on the other hand, behavioral variance is low, and strategic uncertainty is minimal. We investigate both theoretically and empirically how people respond to different levels of strategic uncertainty when they are motivated by social preferences. 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 we explain below, our findings strongly support the latter view.

Using a representative sample of the U.S. population, we examine our research question

[2] 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."

experimentally through the lens of a well-powered (N=1203) and pre-registered study.[3] We do so by introducing a variant of the established public goods game (PGG) with two players as used by Gächter et al. (2017). Participants 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 attitude, as well as their beliefs about their co-player's contribution. In Part II we then present the participants with the distribution from which their 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.

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.

Our findings can be informative for organizations in a variety of ways. They suggest that individuals may react very differently to the same loose or polarized environment. Some will focus on the presence of high contributors, while others will primarily focus on matching the free riders. Accounting for this heterogeneity may be especially important when predicting the effect of a merger, or when considering different strategies to motivate employees.
What determines how an individual reacts to a loose or polarized environment? In this respect, our results point to an interaction between strategic uncertainty and 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). Personal traits such as *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 both personal values and 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 organizations or policymakers who want to nudge a certain type of behavior (see also Barr et al., 2020).

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 about mean

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

or modal behavior in a given situation influences individual decisions. This has been found in both non-strategic settings such as dictator games (Bicchieri and Xiao, 2009; Danilov et al., 2021), voluntary payments (Shang and Croson, 2009) and donations to charities (Dimant, 2019; Bicchieri et al., 2022a), as well as in strategic interactions such as public goods provision (Chen et al., 2010). Consistent with this, we also find that differences in means affect subsequent decisions. However, we extend this literature by considering the whole distribution of behavior, and the resulting degree of strategic uncertainty. In this sense, we relate to the recent literature that looks at the implications of uncertainty about norms for individual behavior (Feldhaus et al., 2019; Fosgaard et al., 2020). In this vein, d'Adda et al. (2020) consider a setup where dictators are shown different distributions of normative views (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. However, the setup they study focuses on normative views and does not feature any direct strategic interaction between participants and the individuals whose normative views are being shown. Instead, our paper focuses on the reaction of participants to the behavior (rather than the normative views) of others in the presence of direct strategic interactions. As mentioned, this implies that different distributions generate different levels of strategic uncertainty. We want to understand how people react to this. The environment we analyze and the question we address are therefore fundamentally different from d'Adda et al. (2020).

Our paper also adds to the literature on heterogeneity in contributions in public good dilemmas. Previous research has documented substantial variation in the distribution of contributions 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. Other studies have investigated how specific conditions of the game itself such as various forms of feedback (see e.g. Chaudhuri et al. (2006); Kerr et al. (2009); Croson (2007); Bigoni and Suetens (2012); Fiala and Suetens (2017) or the heterogeneity in contributions in one's interacting group affects (conditional) cooperation. Using the strategy method, Cheung (2014), Hartig et al. (2015) and Berg et al. (2015) consider setups where individuals interact in groups and the actions of all group members are known. Wolff (2017), finally, elicits conditional contributions in a public good game and computes the Nash-equilibrium sets that result from the participants' elicited preferences. His findings suggest that multiple equilibria are relatively frequent. These papers show that individuals do not react only to averages but to the whole profile of individual contributions. In particular, Berg et al. (2015) find that the participants' (hypothetical) responses to heterogeneity in peer contribution exhibit considerable variation, and are associated with cooperative tendency. Our results extend these findings by providing a systematic investigation of how individuals react to heterogeneity in a context characterized by strategic uncertainty, where co-player's behavior is not known in advance and individuals form expectations based on the descriptive norm.

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, 2023a). 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.[4] Taken together, our research's findings extend beyond the realm of social norms and public goods provision, with implications for firms, organizations, and management practices. Understanding how employees respond to descriptive norms in the workplace can have profound impacts on company culture, productivity, and ultimately, organizational success.

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 additional discussion. Section 6 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; Kölle and Quercia, 2021). 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 substantiates our investigation on the effect of descriptive norms.

We model this concern for reciprocity by letting individuals incur a psychological loss whenever their contribution differs from that of their co-player. This ensures that *ceteris paribus*, individuals adapt their behavior to the behavior of their co-player. This is however not always straightforward. For example, when engaging in everyday interactions, people typically do not know what their counterparts will do. At best, they can form beliefs based on 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 and the risk of contributing too much (relative to their co-player). 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

[4]The software eliciting norm-related beliefs using the distribution builder can be downloaded here.

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, $x_i^a \in [0, \overline{x}]$ captures i's underlying cooperative tendency or what i considers the "right thing to do" (which we call i's *personal value*) and $\delta_i \geq 0$ captures the importance that i ascribes to acting in accordance to their personal value.[5] 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. Note that, in addition to reciprocity concerns, the mismatch loss could equivalently originate from a desire to simply conform to what the co-player is choosing. The specific motivation is irrelevant for the purpose of our results. 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, here individuals are interested in minimizing the average distance between own and co-player's contribution. 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

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

[5] 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.2 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.

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. In the following, we refer to α_i as *sucker* and β_i as *free-riding aversion*.

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. These cumulative distribution functions are stylized illustrations that approximate the cumulative distributions of co-player contributions in our low-variance, high-variance, and u-shaped treatments. The horizontal straight lines represent φ_i.

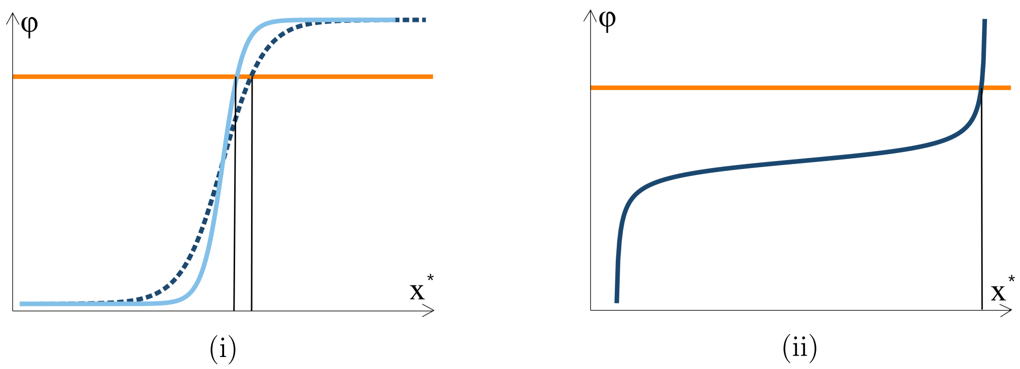


Figure 1: F(x) for single peaked (i) and polarized (ii) distributions

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 (Figure 1 (ii) illustrates the latter possibility). Intuitively, choosing an intermediate contribution tends to be dominated. That's because if x_i is moved marginally the probability that the co-player chooses a higher or a lower contribution remains largely unchanged. 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 vis-à-vis their co-player rather than opting for a "middle of the road" contribution level. 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^a_i, in determining individual contributions. A direct implication of proposition 2 is that contributions increase with free-riding aversion (β_i) and personal values (x^a_i), but decrease with sucker aversion (α_i).

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

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^a_i*) 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)).

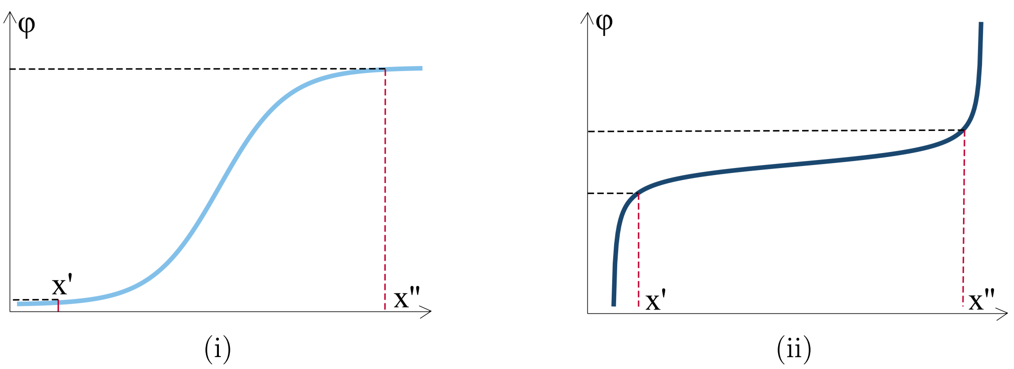


Figure 2

Note that Corollary 3 does not state that individual traits *always* have a larger effect on contributions when individuals face more dispersed descriptive 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 personal traits may have a larger effect when individuals are confronted with more dispersed descriptive norms. This is intuitive: As seen above in Corollary 1, when individuals face a dispersed descriptive norm, their contributions are more dispersed. There is therefore more scope for individual traits to affect 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.

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 Appendix B Figure B.1, 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

To sum up, our theoretical predictions depend on the nature of the mismatch loss in the utility function. In the first case we consider, individuals are primarily concerned with matching their co-player's mean contribution. Whether this mean contribution arises from a tight, loose, or polarized distribution has no bearing on the optimal contribution choice. On the other hand, in the second case we discussed, an individual's optimal contribution also depends on the nature of the distribution of the co-player's contribution. When the distribution of co-player's contribution is more spread out, this affects an individual's contribution by either increasing or decreasing it, depending on the individual's personal traits. Intuitively, some individuals are primarily concerned with avoiding contributing more than their co-player, while others are primarily concerned with avoiding contributing less than their co-player. Once we aggregate all individual choices, this generates a distribution of contribution choices that is also more spread out. We can now lay out the hypotheses that follow from our theoretical investigation. As discussed, there are two alternative hypotheses regarding the impact of the distribution's shape and variance, depending on the assumptions about the underlying loss function.

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, assuming a quadratic loss function, while Hypothesis 1b follows from Proposition 2 and Corollary 1, assuming a loss function based on absolute values. If Hypothesis 1b holds, we can derive further hypotheses for how individual traits affect contributions. As we will describe in greater detail below (see Section 3.3), in our experiment we elicit a measure of both sucker (α_i) and free-riding aversion (β_i). x_i^a, an individual's personal value, is also measured directly within the experiment. From Corollary 2 we expect that contributions increase with x_i^a and α_i, but decrease with β_i.

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

As shown in Corollary 3 and in line with the findings of Elster and Gelfand (2021), we expect the effect of personal traits to differ across different descriptive norms.

Hypothesis 3. *Suppose that Hypothesis 1b 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, in line with Hypothesis 4 and previous experimental evidence, we expect individuals to react to the mean of a given contribution, contributing more the higher the mean.

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

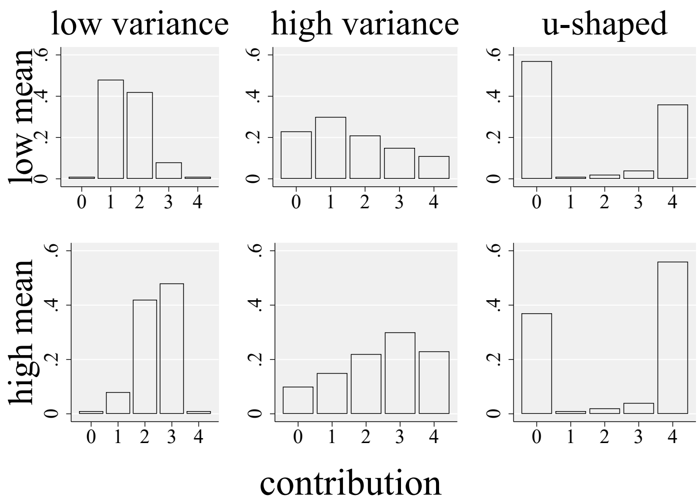
To empirically test the hypotheses derived from our theoretical framework, our experiment exogenously varies the *mean* and *variance/shape* of the co-player's behavior in a two-person PGG. Table 1 gives an overview of the different treatments, including their mean and variance. The corresponding distributions 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 and Gächter, 2010; Gächter et al., 2017) to elicit the underlying cooperative attitudes when individuals interact with a randomly chosen co-player without knowing anything about the underlying distribution of behavior. These attitudes can then be used as controls for analyzing behavior in Part II.

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	1	1.6	1.6	1	1.6	3.6	0
High mean	2.4	0.5	3	2.4	1.6	3	2.4	3.6	4

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



- 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 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. Note that what we call a u-shaped or polarized norm can be equivalently interpreted as reflecting the coexistence of two different norms in the population. Thus, while we use the term "polarized descriptive norm", it should be clear that this is simply a semantic choice and should not be seen as conflicting with a "dual norm" interpretation.

The distributions of co-player behavior in Part II are constructed through non-random sampling from a previous session. Similar approaches have also been used by other studies (see e.g. Frey and Meier, 2004; Bicchieri and Xiao, 2009; Krupka and Weber, 2009; Bursztyn et al., 2020) and the use of non-random samples in experiments is for example discussed by Charness et al. (2022) and Bardsley (2000). 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 was incentive-compatible in that it would affect the payoffs of those previous participants, with one of which they would be paired at random.

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 4 tokens (Ledyard, 1995; Fischbacher et al., 2001). Participants are paid for both their decisions in Parts I and II, but receive no information about their payoffs between parts to reduce potential hedging. 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:

- 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**).

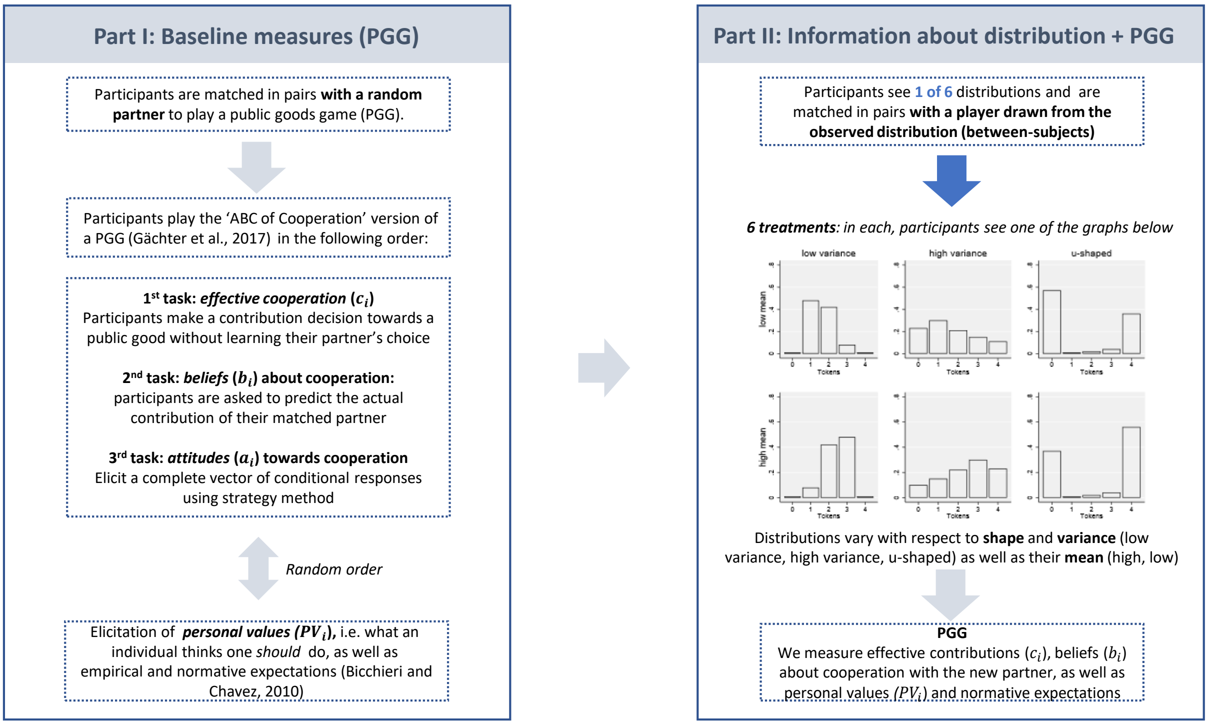
We always elicit unconditional contributions first, followed by beliefs about co-player behavior and conditional contribution attitudes. The elicited attitudes give us a conditional contribution vector that we use to classify participants into different cooperation 'types', namely conditional cooperators, unconditional cooperators, free-riders, triangle cooperators, and others. Subjects are also asked to express their personal values (PVs). In particular, we ask participants what in their opinion is the most appropriate amount to invest, where appropriate means "correct" or "moral" (Bašić and Verrina, 2020). Measuring personal values offers additional insights into behavior in our setting, as they provide an indication of an individual's personal norm or unconditional cooperative tendency (Catola et al., 2021). They are also widely used as a part of norm elicitation procedures in the literature (Bicchieri and Chavez, 2010; Bašić and Verrina, 2020).[6]

In line with our research question, we measure participants' beliefs as a whole distribution. To do so, we follow the approach discussed in Dimant (2023a) and ask participants to allocate points across all possible co-player contributions (see Appendix B, Figure B.2).The more likely participants think a contribution is, the more points they should allocate to it. To incentivize decisions, we use a quadratic scoring rule adapted from Artinger et al. (2010), coupled with an intuitive visual interface (see Quentin, 2016). The quadratic scoring rule is communicated to participants in a way that separates gains from correctly assigned points and losses from incorrectly assigned points. Artinger et al. (2010) show that this way of communication improves understanding (see Appendix C for more details on the used scoring rule). This elicitation method allows us to obtain a sense of the participant's beliefs about the whole distribution of co-player contribution in both Part I and Part II of the experiment. In addition to the mean of the distribution that we obtain from each participant (which we refer to as average individual belief), this allows us to calculate the standard deviation of an individual's beliefs as a measure of dispersion in our analysis (SD of individual beliefs).

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 decisions participants take in the PGG in Part II have real consequences. They determine the size of the bonus for participants from the previous sessions used to construct the distributions, as well as their own bonus. Figure 4 gives an overview of the design.

[6] 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.3), these are very similar to personal values and beliefs about the co-player's behavior.

Figure 4: Overview of the experimental design



3.3 Sample and data collection

The experiment and our hypotheses were pre-registered (see here) in November 2021. We programmed the experiment using Qualtrics (2005) and recruited participants online via Prolific in December 2021. In total, we recruited a sample of about 1,200 US participants who are representative in terms of age, gender, and ethnicity, resulting in ∼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 an additional bonus depending on the decisions of the participants in the main experiment.

After the main experiment, participants completed an ex-post survey that provided 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 (see Figure B.4 in Appendix B for manipulation checks). As we told participants that the distribution from which their co-player's contribution was drawn was taken from one of the six subgroups we constructed from previous sessions, we also asked how common they thought this behavior was. Finally, we used two questions to proxy the participants' aversion towards contributing more and less than their co-player. As a measure of sucker aversion, we asked participants "*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?*". To measure free-riding aversion, we asked "*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?*"

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 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*. The distribution of types is independent of our treatments (χ^2-test, $p = 0.35$).

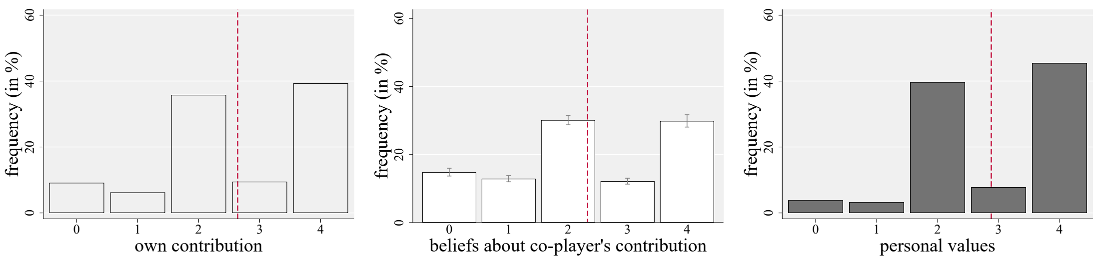
Figure 5 shows contributions, aggregate expected co-player contribution, and personal values in Part I. The most frequent contribution levels are 2 and 4 tokens, revealing relatively high levels of cooperation. The same is true for beliefs about the other player's contribution and personal values about what is the "right thing to do" in the game.[7]

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. To do so, we split participants into two groups - those with a variance above the median and those with a variance below the median. We then compare unconditional contributions (the C component in the ABC elicitation) in both groups. We find that, in the sample of participants with a high variance of beliefs, the distribution of unconditional contributions exhibits greater variance (F-test, p = 0.001). While this analysis cannot provide causal evidence, it gives initial anecdotal support for the notion that variance matters and that looser environments may generate more varied responses. We also find that participants who expect their co-player to contribute more also tend to contribute more themselves (r=0.65, $p < 0.001$). 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 partic-

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.

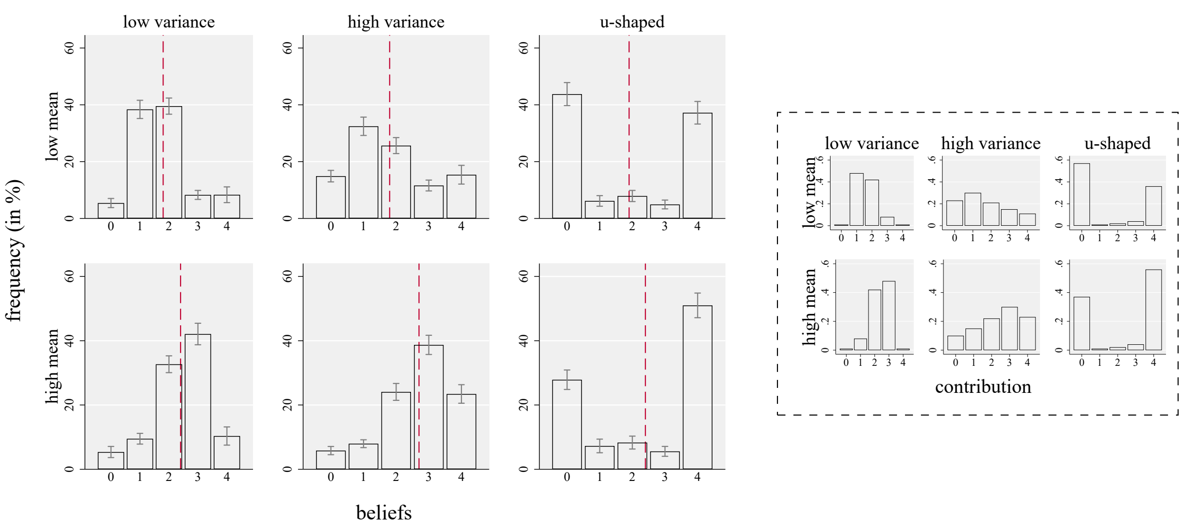
[7]In Appendix B (Figures B.5 and B.6) we provide a more detailed analysis on belief heterogeneity.

ipants' 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 (see Figure B.7 in Appendix B for changes in beliefs between parts).

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 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. Finally, Figure 7 shows that the polarized norm induces polarization in subsequent contribution behavior ($\sigma^2 = 2.68$). In Section 5.2 we provide a discussion of the personal traits that drive this heterogeneity.

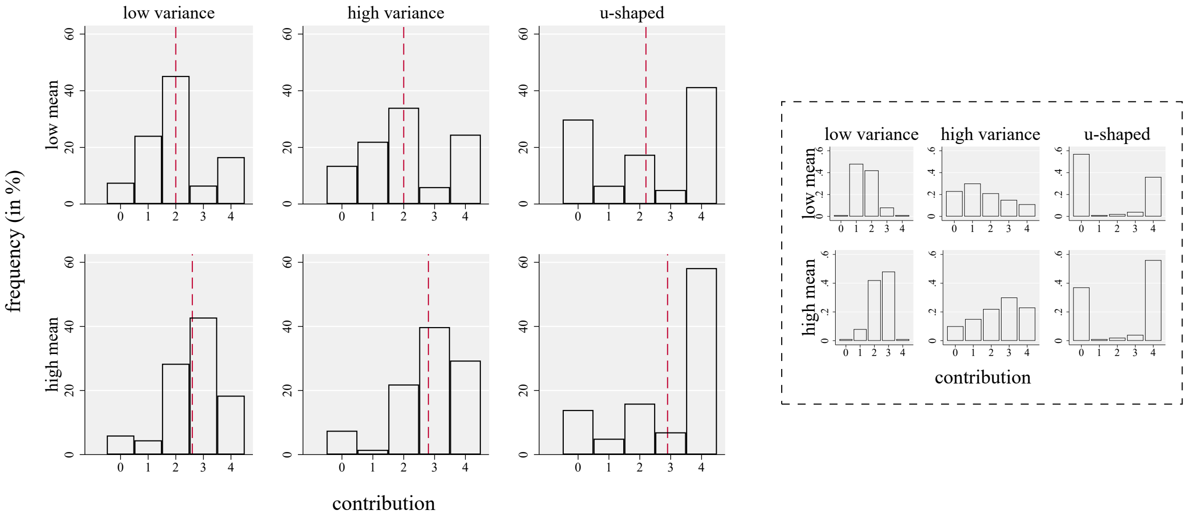
To test the first visual impression of treatment effects, we perform pairwise F-tests for the equality of standard deviations between treatments. Overall, we find that the variance in contributions is significantly higher in polarized than in tight or loose environments (for both $F < 0.001$). Loose environments in turn generate a significantly higher variance in participants' contributions than tight environments ($F = 0.007$). Pairwise χ^2 tests 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 ($p < 0.001$). Figure B.8 shows in more detail how individuals change contributions across parts. As a further test of treatment differences, we consider two subgroups, high contributors (≥ 2) and low contributors (≤ 2), and show that, in treatments characterized by greater variance, high contributors contribute significantly more, while low contributors contribute significantly less, thus generating higher variance of contribu-

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.

tions overall (see Table B.1, Appendix B).[8] 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 and lead to further polarization.*

4.3 Personal traits

We now look at the effect of personal values, free-riding, and sucker aversion on contributions. In Table 2 we regress contributions on these personal traits, treatment indicators for the variance/shape of the observed distribution and their interaction. 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.[9] 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

[8]These results are robust to controlling for how common participants perceive the shown distribution to be (Appendix B, Table B.2) and to using beliefs in Part II instead of treatment indicators as regressors (Table B.3).
[9]Alternative estimation methods such as OLS or ordered probits yield qualitatively similar results.

Table 2: Tobit models. Effect of personal traits on contributions in Part II

	(1)	(2)
High mean	0.96***	0.97***
	(0.12)	(0.12)
Variance (baseline = low)		
High variance	-0.51	-0.59
	(0.52)	(0.51)
U-shaped	-0.41	-0.51
	(0.55)	(0.54)
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. Results of two tobit regressions (data is censored at 0 and 4). The dependent variable is *contributions in Part II*. *Personal values* are measured in Part I and can take values between 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. *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 GPS measures for negative and positive reciprocity. All regressions control for order effects.

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 environment and sucker aversion, as well as between u-shaped environment and free-riding aversion are highly significant in the expected direction. The interaction with personal values is both significant for the high-variance and the u-shaped environment, confirming that personal values have a stronger effect on decisions in these conditions (see Hypothesis 3).[10]

[10] We use personal values in Part I as these are collected “in a vacuum” and are therefore not affected by the treatments. Our results are robust to controlling for changes in personal values between the two parts of the experiment (which are very minor) and their interaction with treatment indicators (see Appendix B, Table B.4).

Result 3. *Personal traits (sucker-version, free-riding aversion, and personal values) matter more when participants are confronted with loose or polarized descriptive norms than in the presence of tight descriptive norms.*

Result 3 highlights the role of personal traits for individual contributions when strategic uncertainty is high. Intuitively, when strategic uncertainty is high participants face a lot of uncertainty about their co-player's contribution. Since matching their co-player's contribution is not straightforward, participants rely on their personal preferences to guide their contribution choice. Personal traits are further discussed in Section 5.2.

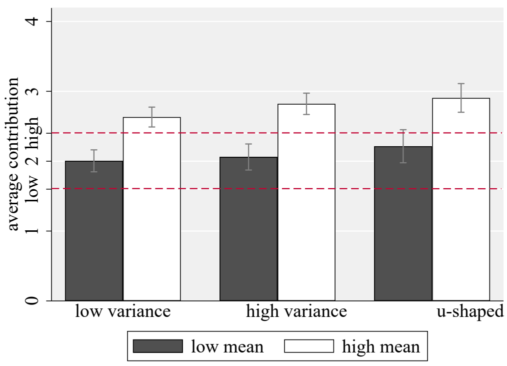
4.4 Effect of high and low means

Finally, we look at the difference between descriptive norms with high and low means (see Hypothesis 4). 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 of the observed distribution (see Figure 8).

Table 3 tests Hypothesis 4 formally, by regressing contributions on treatment indicators for the mean and variance/shape of the observed distribution. Models 4–6 moreover include an interaction term between mean and variance/shape. As expected, contributions are significantly higher in the high-mean conditions. This finding also holds when controlling for baseline behavior and including demographic controls. 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 selfish behavior. However, our data suggest that this is not the case. Greater strategic uncertainty promotes on average higher contributions. 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. Finally, models 4–6 show that the effect of the mean seems to be even larger in the polarized condition, as indicated by the significant interaction.

Result 4. *Individuals 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.

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***	0.94***	0.70***	0.63***	0.57***
	(0.14)	(0.11)	(0.10)	(0.23)	(0.18)	(0.17)
Variance (baseline = low)						
High variance	0.23	0.26**	0.27**	0.08	0.09	0.08
	(0.16)	(0.13)	(0.12)	(0.23)	(0.18)	(0.17)
u-shaped	0.61***	0.65***	0.60***	0.29	0.25	0.22
	(0.17)	(0.13)	(0.13)	(0.23)	(0.18)	(0.18)
Interactions						
High mean x high variance				0.29	0.35	0.37
				(0.33)	(0.26)	(0.25)
High mean x u-shaped				0.65*	0.82***	0.80***
				(0.34)	(0.27)	(0.26)
Constant	2.05***	-1.31***	-1.95***	2.21***	-1.14***	-1.74***
	(0.16)	(0.22)	(0.58)	(0.19)	(0.23)	(0.59)
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	0.15	0.17

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.
Note. Results of Tobit regressions (data is censored at 0 and 4). The dependent variable is *contributions in Part II*. *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, average individual beliefs, SD of individual 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 GPS measures for negative and positive reciprocity. All regressions control for order effects.

Before providing a wider discussion on our results, note that as is universally true for experimental research, the existence and role of experimenter demand effects (EDEs) should be considered (Zizzo, 2010). In our setup, we are not concerned with the potential presence of EDEs, for two reasons. First, in Part II of the experiment, each participant makes a single contribution choice. The finding that loose norms lead to loose responses and polarized norms to polarized ones arises from aggregating these unique individual choices in each treatment. It does *not* rely on each individual participant generating a distribution of contributions. Put differently, the result arises because in loose or polarized environments different people react very differently, while in tight environments almost everyone reacts in a similar fashion. 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.

5 Additional analysis and discussion

5.1 Exposure to strategic risk

An interesting feature of our findings is that a large share of participants make choices that expose them to considerable strategic risk. Consider, for instance, a situation where a share $p < 1/2$ of the population contributes the maximum of 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.[11] 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 conditionally reciprocal and thus exhibit concern for matching their co-player's contribution, they are not mismatch-risk-averse. This suggests that in 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 (see e.g. Chetty (2006) and more generally the vast literature on monetary risk aversion).

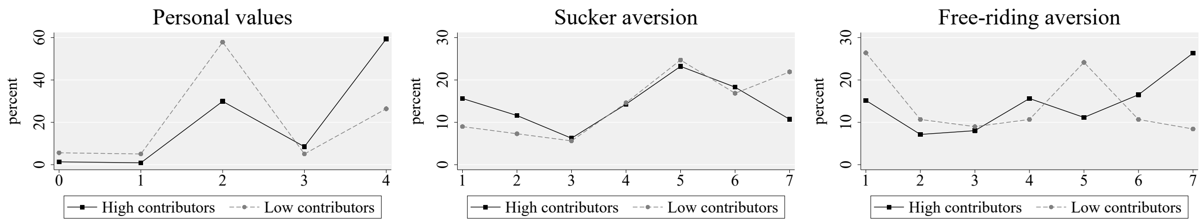
5.2 The role of personal traits

As discussed in Section 5.2, personal traits influence participants' responses to environments with high strategic uncertainty. This is why we see a wider range of contributions in these environments compared to tight environments, where strategic uncertainty is low. To delve into this further, let's re-examine a polarized environment where strategic uncertainty is at its highest and participants are aware that their co-player is most likely either a full contributor or a free-rider. Our results imply that participants' choice of contribution — whether they focus more on the likelihood of full contribution or on the possibility of free-riding by their co-player — is shaped by personal traits

To back up our arguments, Figure 9 reproduces the distributions of personal traits of those participants faced with a u-shaped distribution who choose to contribute more than 2 vs those who contribute 2 or less. As Figure 9 indicates, the distributions of personal traits differ between high and low contributors. In particular, low contributors are considerably more likely to exhibit extreme sucker aversion while high contributors are considerably more likely to exhibit extreme free-riding aversion. Personal values also tend to be lower among low contributors. This visual impression is confirmed when looking at average values, which are statistically different across the two subgroups for all personal traits (Wilcoxon signed-rank tests, $p < 0.001$). Interestingly, we also find equivalent differences in the low and high-variance treatments. In fact, across treatments, the distributions of personal traits in each subgroup are statistically indistinguishable (see Appendix B.7). The same types of people choose high or low contributions in all our treatments. However, as we have seen, the contribution choices within each subgroup *do* differ between treatments. High contributors are most generous in the u-shaped treatment, followed by the high-variance treatment, while the opposite happens for those in the low contributors subgroup. In Appendix B, Table B.5, 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 show lower levels of both types of aversion. Finally,

[11] 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 $p < 1/2$.

Figure 9: Distribution of personal traits among low & high contributors in the u-shaped condition



free-riding aversion and sucker aversion are positively correlated at the individual level ($r = 0.37, p < 0.001$) and are higher for conditional cooperators compared to other types (Wilcoxon-Mann-Whitney test, $p < 0.001$).[12]

5.3 Contributions in Part I

The analysis of Part II focuses on elicited personal values as a measure of unconditional cooperative tendency. An alternative is to use unconditional contributions in Part I (elicited in Part C of the ABC protocol), in line with the approach taken by previous literature such as Berg et al. (2015). Because personal values and contribution in Part I are strongly correlated ($r = 0.64$, $p < 0.001$), controlling for both in our regression analysis would be problematic. If we replace personal values with Part I contributions in the regression presented in Table 2, we find that the qualitative message is unchanged. As shown in table B.6 in Appendix B, the interaction effect between Part I contribution and high-variance or u-shaped is positive and significant. This is not surprising given that both personal values and Part I contributions reflect an individual's general cooperative tendency. The effect of a high cooperative tendency on own contribution is stronger when the descriptive norm is loose or polarized.

5.4 The role of strategic interactions

Strategic interactions are pervasive in everyday life as well as in firms and organizations. This makes them especially interesting and important to study. Having said that, we believe that our findings may potentially apply more widely. Consider for example a non-strategic environment (say, a charity donation) where individuals want to conform to their empirical expectations, i.e. they want to conform to what they believe other individuals do. Although this setup does not feature *direct* strategic interactions, these may arise *indirectly* from the agents' desire to conform. Formally, suppose that i) wanting to conform to empirical expectations takes the form of (or is isomorphic to) not wanting to deviate too much from the action of a randomly drawn individual as summarized e.g. by a psychological cost $\int L(a_i, a_j) f(a_j)\, da_j$, where $f(.)$ characterizes empirical expectations and $L(.)$ is a loss function increasing in the difference between the two actions,

[12] Sucker and free-riding aversion also correlate with an individual's contribution choices in the A component of the ABC elicitation method. Sucker aversion is negatively correlated with a participant's conditional contribution in the event that the co-player contributes zero ($r = -0.19, p < 0.001$). Conversely, free-riding aversion is positively correlated with a participant's conditional contribution in the event that the co-player contributes everything ($r = 0.24, p < 0.001$).

and ii) the mismatch loss is proportional to the absolute value of the difference between the two actions in a manner akin to expression (2). This might occur for instance if people care about how their behavior compares to that of another randomly drawn individual. They don't want to look like "suckers", but also don't want to appear too selfish. Then, the insights and empirical predictions of our model apply directly. Intuitively, that's because different distributions of behavior affect the strategic uncertainty faced by individuals. In fact, the theory predicts that people will react to changes in the variance of the descriptive norm as they do in our setup. Hence, while (direct) strategic interactions provide a natural environment to study strategic uncertainty, it is important to stress that these are not *necessary* for our mechanism to apply.

5.5 Long-run dynamics

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. A possible conjecture is that in a dynamic setting, individual traits may exhibit some form of history dependence. For instance, if an individual takes a risk and behaves cooperatively, but is disappointed by the person they are interacting with, this could induce them to behave less cooperatively in the future, even if confronted with the same distribution of behavior. Similarly, if an individual contributes little but is pleasantly surprised by the person they are matched with, they could feel remorseful and behave more cooperatively in future interactions. Formally, this would require that sucker- and free-riding aversion may be affected in a non-negligible way by experiences in the previous period. This strong form of history dependence would generate a very complex system where the distribution of contributions (and thus the descriptive norm) is constantly evolving. The question is then whether the system would converge to a stable distribution over time, and, if yes, whether this can be characterized as a function of initial conditions. Ultimately, the answer to this question is empirical and should be investigated in future work.

6 Concluding remarks

Organizations are increasingly using information about descriptive norms to nudge positive behavior change. Whether it is to increase safety, civility, or ethical behavior, providing information about what most people do in organizational settings can nudge people to behave differently in a wide range of important domains (for a review, see e.g. Venema and van Gestel, 2021). In this study, we investigate how different descriptive norms of cooperative behavior affect an individual's own willingness to cooperate. While previous research has focused almost exclusively on average behavior, we argue both theoretically and empirically that not only averages but also the *shape* of the whole distribution of behavior is an important determinant of reactions to descriptive norms. 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 may 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 support 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".

Another key finding is that an individual's reaction to high strategic uncertainty is moderated by their personal traits. Relative to environments where strategic uncertainty is low, decisions under high uncertainty are more strongly influenced by personal traits such as personal values, i.e. what an individual thinks is the right thing to do in this situation.

This has practical implications for organizations and policymakers interested in 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, managers may need to focus on personal values to be more successful, whereas when intervening in contexts with tight norms, it may be better to focus on the behavior of others. Our findings on the role of descriptive norms, their variance, and shape in influencing behavior, are thus particularly relevant for change management and organizational culture shifts. Managers and leaders can leverage the understanding of how norms influence behavior when implementing changes or attempting to shift the prevailing organizational culture. For instance, when attempting to change from a loose culture (high variance in behavior) to a more cohesive one (low-variance in behavior), leaders can provide employees with descriptive norms that reflect the desired behavior distribution. This could be accomplished through clear communication, training, and role modeling (Cialdini and Goldstein, 2004). The implications of strategic uncertainty can also be used to align individual motivations with organizational goals and foster a positive work environment. If the organization wishes to discourage free-riding and encourage full contribution from its employees, then openly communicating expected performance levels and providing a clear view of the spread of contribution levels across the workforce may be an effective strategy (Vigoda-Gadot, 2007).

Our results provide many other avenues for future research. For example, existing literature suggests that punishment often plays an important role in 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 exposure to tight, loose, and polarized descriptive norms generates different responses even in the *absence* of a norm enforcement mechanism. Our study thus represents a conservative test of the underlying mechanism. Future research could investigate the role of different behavioral patterns in a setup that incorporates punishment. For instance, 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 interesting direction for further research is to explore how people navigate risk in the social domain. As discussed, in this paper we find that in polarized environments a large share of participants choose the maximum contribution, thus exposing themselves to considerable strategic risk. 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. In future research, it would be interesting to explore this idea further. Finally, our work could be extended by moving beyond WEIRD (Western, Educated, Industrialized, Rich, Democratic, see Henrich et al., 2010) samples to test the generalizability of our results.

Overall, we show in this paper 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 organizations and societies. The nuances of how individuals react to different types of descriptive norms, as uncovered by our research, moreover provide valuable insights into how organizations can manage and influence employee behavior, particularly in environments marked by strategic uncertainty. The strategic implementation of these insights could yield substantial improvements in organizational culture, productivity, and overall success. 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.

Ai, W., Chen, Y., Mei, Q., Ye, J., and Zhang, L. (2023). Putting teams into the gig economy: A field experiment at a ride-sharing platform. *Management Science*.

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*.

Aycinena, D., Rentschler, L., Beranek, B., and Schulz, J. F. (2022). Social norms and dishonesty across societies. *Proceedings of the National Academy of Sciences*, 119(31).

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. *CeDEx Discussion Paper Series*, (2020-06).

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. In Newman, P., editor, *The New Palgrave Dictionary of Economics and the Law*. Palgrave Macmillan London.

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

Belle, N. and Cantarelli, P. (2021). Nudging public employees through descriptive social norms in healthcare organizations. *Public Administration Review*, 81(4):589–598.

Berg, P. v. d., Molleman, L., Junikka, J., Puurtinen, M., and Weissing, F. J. (2015). Human cooperation in groups: variation begets variation. *Scientific Reports*, 5(1):1–7.

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. and Dimant, E. (2023). Norm-nudging: Harnessing social expectations for behavior change. Working Paper.

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.

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

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.

Burks, S. V. and Krupka, E. L. (2012). A multimethod approach to identifying norms and normative expectations within a corporate hierarchy: Evidence from the financial services industry. *Management Science*, 58(1):203–217.

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

Burtch, G., Hong, Y., Bapna, R., and Griskevicius, V. (2018). Stimulating online reviews by combining financial incentives and social norms. *Management Science*, 64(5):2065–2082.

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.

Catola, M., D'Alessandro, S., Guarnieri, P., and Pizziol, V. (2021). Personal norms in the online public good game. *Economics Letters*, 207(110024):110024.

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. and Goldstein, N. J. (2004). Social influence: Compliance and conformity. *Annu. Rev. Psychol.*, 55:591–621.

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.

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

Cullen, Z. and Perez-Truglia, R. (2023). The salary taboo privacy norms and the diffusion of information. *Journal of Public Economics*, 222:104890.

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.

Danilov, A., Khalmetski, K., and Sliwka, D. (2021). Descriptive norms and guilt aversion. *Journal of Economic Behavior & Organization*, 191:293–311.

Danilov, A. and Sliwka, D. (2017). Can contracts signal social norms? experimental evidence. *Management Science*, 63(2):459–476.

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. (2023a). Beyond average: A method for measuring the tightness, looseness, and polarization of social norms. *Economics Letters*.

Dimant, E. (2023b). 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.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., and Wagner, G. G. (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the european economic association*, 9(3):522–550.

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.

Falk, A., Becker, A., Dohmen, T., Huffman, D., and Sunde, U. (2023). The preference survey module: A validated instrument for measuring risk, time, and social preferences. *Management Science*, 69(4):1935–1950.

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.

Fiala, L. and Suetens, S. (2017). Transparency and cooperation in repeated dilemma games: a meta study. *Experimental Economics*, 20:755–771.

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.

Fosgaard, T. R., Gårn Hansen, L., and Wengström, E. (2020). Norm compliance in an uncertain world. *IFRO Working Paper.*

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 mismatches in response to collective threat. *Current Directions in Psychological Science*, 30(5):401–409.

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.

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.

Guiso, L., Sapienza, P., and Zingales, L. (2015). Corporate culture, societal culture, and institutions. *American Economic Review*, 105(5):336–339.

Haerpfer, C., Inglehart, R., Moreno, A., Welzel, C., Kizilova, K., Diez-Medrano, J., Lagos, M., Norris, P., Ponarin, E., and Puranen, B. (2022). World values survey wave 7 (2017-2022) cross-national data-set. *World Values Survey Association*.

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.

Huck, S., Kübler, D., and Weibull, J. (2012). Social norms and economic incentives in firms. *Journal of Economic Behavior & Organization*, 83(2):173–185.

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.

Jacobson, R. P., Marchiondo, L. A., Jacobson, K. J., and Hood, J. N. (2020). The synergistic effect of descriptive and injunctive norm perceptions on counterproductive work behaviors. *Journal of Business Ethics*, 162:191–209.

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.

Kimbrough, E. O., Krupka, E. L., Kumar, R., Murray, J., Ramalingam, A., Sánchez-Franco, S., Sarmiento, O. L., Kee, F., and Hunter, R. (2022). On the stability of norms and norm-following propensity: A cross-cultural panel study with adolescents. *Available at SSRN*.

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

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.

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.

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.

Pek, S., Turner, N., Tucker, S., Kelloway, E. K., and Morrish, J. (2017). Injunctive safety norms, young worker risk-taking behaviors, and workplace injuries. *Accident Analysis & Prevention*, 106:202–210.

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*.

Robertson, J. L. and Barling, J. (2013). Greening organizations through leaders' influence on employees' pro-environmental behaviors. *Journal of Organizational Behavior*, 34(2):176–194.

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.

Schultz, W. P., Khazian, A. M., and Zaleski, A. C. (2008). Using normative social influence to promote conservation among hotel guests. *Social Influence*, 3(1):4–23.

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.

Sliwka, D. (2007). Trust as a signal of a social norm and the hidden costs of incentive schemes. *American Economic Review*, 97(3):999–1012.

Stephan, U. and Uhlaner, L. M. (2010). Performance-based vs socially supportive culture: A cross-national study of descriptive norms and entrepreneurship. *Journal of International Business Studies*, 41:1347–1364.

Taylor, S. G., Locklear, L. R., Kluemper, D. H., and Lu, X. (2022). Beyond targets and instigators: Examining workplace incivility in dyads and the moderating role of perceived incivility norms. *Journal of Applied Psychology*, 107(8):1288.

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.

Venema, T. and van Gestel, L. (2021). Nudging in the workplace: Facilitating desirable behaviour by changing the environment. In *A handbook of theories on designing alignment between people and the office environment*, pages 222–235. Routledge.

Vigoda-Gadot, E. (2007). Leadership style, organizational politics, and employees' performance: An empirical examination of two competing models. *Personnel Review*, 36(5):661–683.

Weber, R. A. and Camerer, C. F. (2003). Cultural conflict and merger failure: An experimental approach. *Management Science*, 49(4):400–415.

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

Xie, J. L. and Johns, G. (2000). Interactive effects of absence culture salience and group cohesiveness: A multi-level and cross-level analysis of work absenteeism in the chinese context. *Journal of Occupational and Organizational Psychology*, 73(1):31–52.

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

A.1 Proofs

Proof of proposition 1 Expected payoff is

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

which can be rewritten as

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

The derivative of (3) with respect to x_i gives

$$-(1-\gamma) - \eta_i(x_i - \mu) - \delta_i (x_i - x_i^a) . \quad (4)$$

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:

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

which can be rewritten as

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

The derivative of (5) with respect to x_i gives

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

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

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

■

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 $-(1-\gamma) - \alpha_i F(x) + \beta_i(1 - F(x)) - \delta_i(x - x^a_i)$ 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 $F_1(x') > F_0(x')$ while $F_1(x'') < F_0(x'')$. 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 $F_1(x'') < F_0(x'')$,

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

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.2 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 $F_1^{-1}(x'') > F_0^{-1}(x'')$,

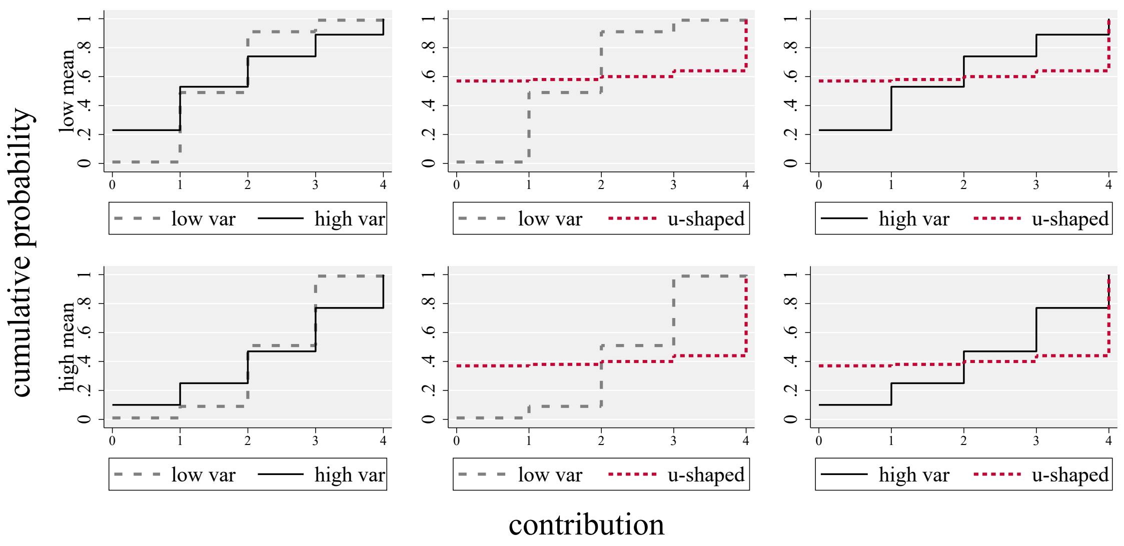
$\widetilde{x}'' - \widetilde{x}'$ is larger under f_1 compared to f_0. We conclude that, if the disutility of deviating from x_i^a takes the absolute value form, corollary 3 does not necessarily hold for x_i^a (but it does for α_i and β_i).

B Additional analysis

B.1 Single crossing property of shown distributions

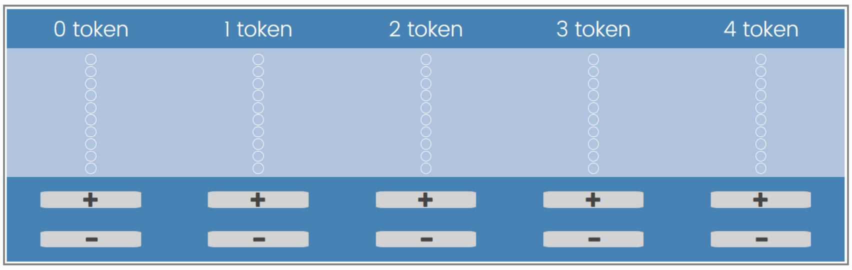
Figure B.1 plots the cumulative distributions of all six treatments. As can be seen below, the single-crossing condition discussed in Section 2 holds for all pairwise comparisons of descriptive norms sharing the same mean.

Figure B.1: Cumulative distributions of descriptive norm treatments



B.2 Belief elicitation screen

Figure B.2: Belief elicitation screen

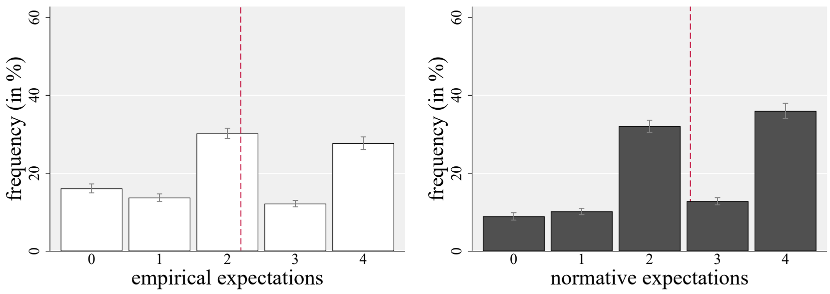


Note. Participants are asked to allocate a total of 10 points across all available options, according to how likely they think each option is.

B.3 Empirical and normative expectations

Empirical (EE) and normative (NE) expectations in Part I — i.e. expectations about what most people do and what most people think one should do — are very similar to contributions, beliefs and PVs (see Figure B.3). 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 measured the participants' confidence in their replies. Participantsappear 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.3: EEs and NEs in Part I

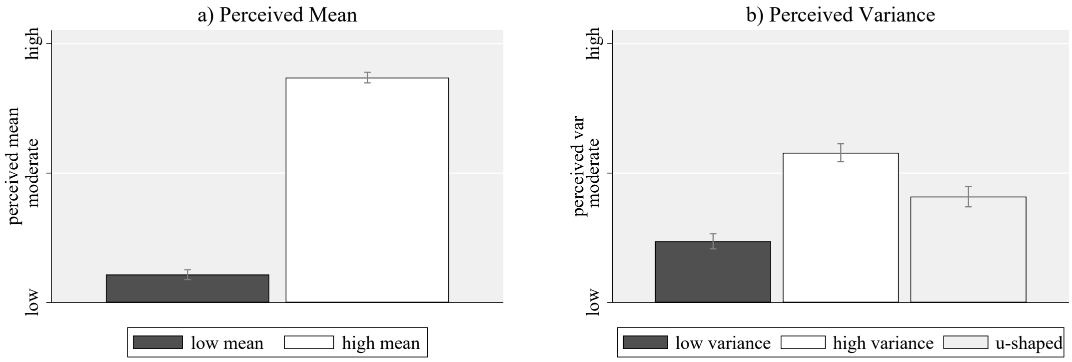


Note. The dashed lines represent averages. Whiskers show 95% confidence intervals.

B.4 Manipulation checks

In the ex-post survey we ask participants about their perceptions of the shown distribution. As Figure B.4a 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.4b), 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-variance 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.4: Manipulation checks

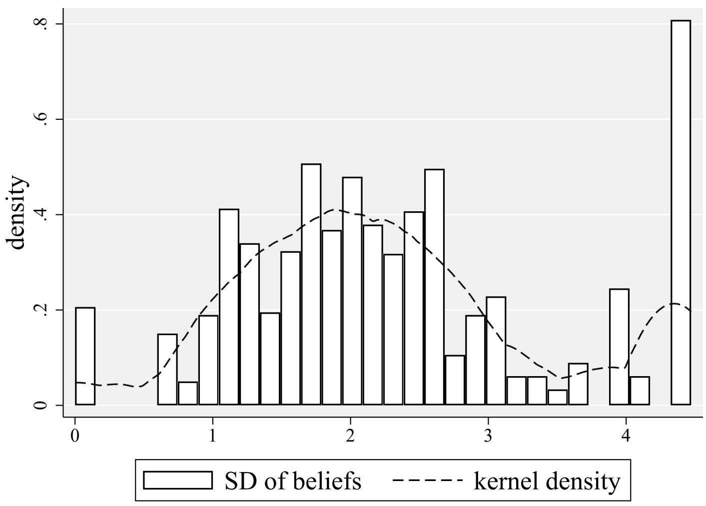


Note. Whiskers show 95% confidence intervals.

B.5 Distribution of beliefs in Part I

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.5 shows that the standard deviation of initial beliefs varies substantially between participants (min=0, max=4.5).

Figure B.5: Variation in standard deviations of beliefs in Part I



The individual distributions of beliefs can roughly be categorized 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

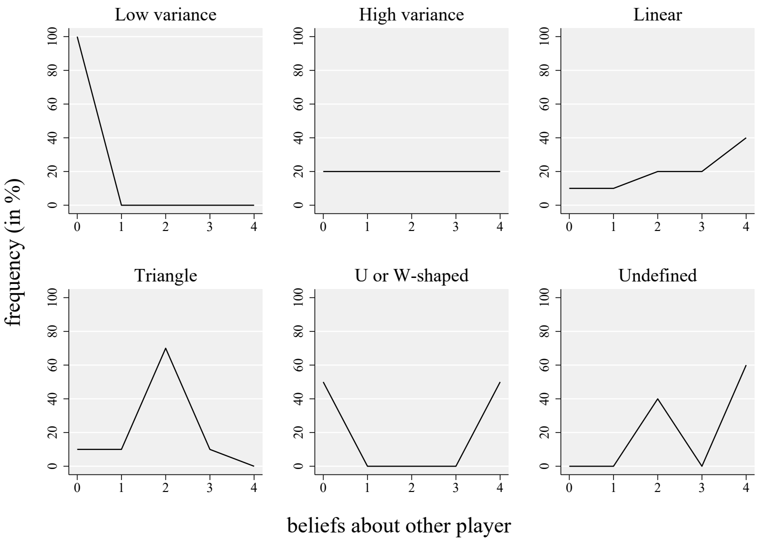
4. Triangle: participants with a single modal belief below 80% that is 1,2, or 3 tokens

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

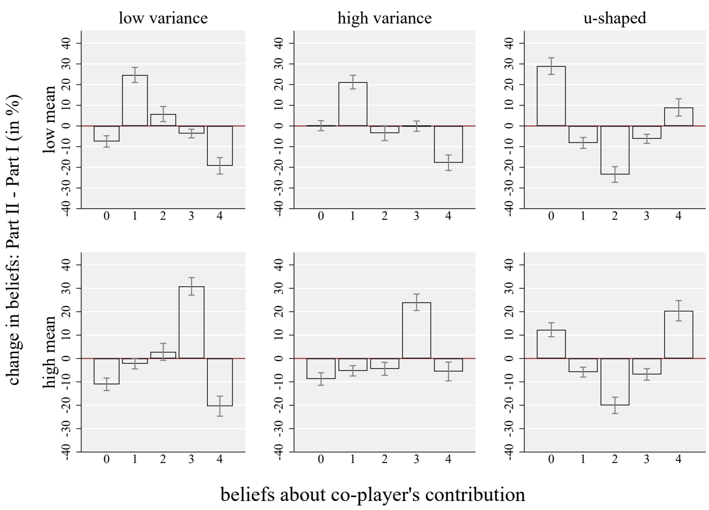
Figure B.6 provides examples for each type. Using these rules we can classify 92% of participants. The most common types are triangles (41%), followed by low-variance (21%), u-/w-shaped (15%), linear (10%), and high-variance types (5%).

Figure B.6: Examples of individual distributions of beliefs in Part I



B.6 Changes between Part I and II

Figure B.7: Changes in beliefs about the co-player's distribution between parts



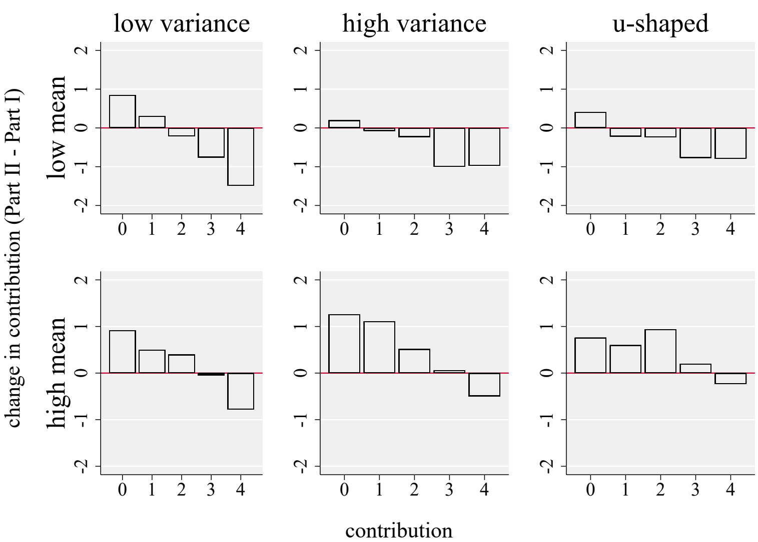
Note. Whiskers show 95% confidence intervals.

Figure B.7 shows changes in beliefs about their co-player's contribution between Part I and Part II. As we can see, 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, $p < 0.001$).

Our data allows us to analyze in more detail how different types of individuals — based on their Part I behavior — respond to the shown distribution in Part II. Figure B.8 shows the change in contributions between Part I and Part II depending on the unconditional contribution level in Part I. Even though individuals with a middle contribution in Part I change slightly less, there is no significant difference in changes across the contribution levels in Part I. When zooming in on the u-shaped conditions, we find that individuals who contributed 0 tokens in Part I are significantly less likely to contribute fully and significantly more likely to contribute 0 in Part II compared to those with a middle contribution (2) in Part I (tobit regression, $p < 0.001$). Conversely, participants who contributed 4 tokens in Part I are significantly more likely to contribute fully and significantly less likely to contribute 0 in Part II compared to those with a middle contribution (tobit regression, $p < 0.001$ and $p = 0.02$ respectively).

Finally, Figure B.8 shows some heterogeneity in how individuals change their behavior between Part I and II across treatments. For low-mean treatments there is a stronger adjustment of high contribution levels, while the opposite holds true for low contribution levels. We also find that individuals change their contributions more in low-mean treatments (Wilcoxon-Mann-Whitney test, $p = 0.03$). Finally, changes are more pronounced in the low-variance (Wilcoxon-Mann-Whitney test, $p = 0.04$) and the u-shaped (Wilcoxon-Mann-Whitney test, $p = 0.08$)

Figure B.8: Changes in behavior between parts by initial contribution level



compared to the high-variance treatment.

B.7 Effect of variance and shape of the distribution

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.

Before discussing our analysis, we note that the distributions of personal traits – sucker aversion, free riding aversion and personal values – in the high (resp., low) contributors sub-samples are statistically indistinguishable across treatments (pairwise Kolmogorov-Smirnov tests). This is also true when we exclude the medium contribution level of 2 from the two subgroups and

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

	(1) High contributions ($\geq$2)	(2) Low contributions ($\leq$2)
Variance (baseline = low)		
High variance	0.32***	-0.13
	(0.09)	(0.09)
U-shaped	1.20***	-0.79***
	(0.10)	(0.10)
High mean	0.36***	0.35***
	(0.07)	(0.08)
Constant	0.62	0.22
	(0.43)	(0.43)
Baseline controls	Yes	Yes
Demographic controls	Yes	Yes
N observations	906	602
Pseudo R^2	0.20	0.14

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

Note. Data is censored at 0 and 4. The dependent variable is *contributions in Part II*. *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 average individual beliefs, SD of individual 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 GPS 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). Regressions control for order effects.

reassures us about subgroups cross-treatment comparability.[13]

Table B.1 shows the results of two tobit models, where we regress contributions in Part II on treatment conditions for each sub-sample. We can see that in fact in the high contribution sub-sample, high-variance significantly increases contributions compared to low-variance, and u-shaped considerably increases contributions compared to high-variance. 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 conditions increase the overall variation of contributions.

As an additional robustness check, we consider the participants' perception of how common is the distribution from which their co-player's contribution is drawn. This allows to control for the possibility that some distributions may be considered more "natural" than others which in principle may act as a confounding effect. In the questionnaire we asked participants how

Table B.2: Tobit models. Effect of variance in observed distribution on contributions in Part II for different sub-samples - controlling for perception of distribution

	(1) High contributions (≥2)	(2) Low contributions (≤2)
Variance (baseline = low)		
High variance	0.33***	-0.13
	(0.09)	(0.09)
U-shaped	1.20***	-0.78***
	(0.10)	(0.10)
Distribution perceived as common	0.06	0.06
	(0.08)	(0.09)
High mean	0.35***	0.35***
	(0.07)	(0.08)
Constant	0.60	0.21
	(0.43)	(0.43)
Baseline controls	Yes	Yes
Demographic controls	Yes	Yes
N observations	906	602
Pseudo R^2	0.20	0.14

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

Note. Data is censored at 0 and 4. The dependent variable is *contributions in Part II*. *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. *Distribution perceived as common* is a binary variable that takes the value of 1 if participants state they rather think the observed distribution is common and 0 otherwise. Baseline controls include average individual beliefs, SD of individual beliefs, and PVs in Part I and can take values between 0 and 4. 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). Regressions control for order effects.

[13] We also compare average personal traits across treatments within each subgroup. (i) There are no significant cross-treatment differences in average free riding aversion; (ii) There is no significant difference in average sucker aversion among low contributors. Among high contributors average sucker aversion is higher in the high-variance than in the u-shaped treatment (p=0.09). If anything, this should reduce the high subgroup's contributions in the u-shaped compared to the high-variance treatment, which goes against what we are looking for; (iii) There are no significant difference in average personal values among low contributors. Among high contributors personal values are higher in the low-variance treatment than in the u-shaped (p=0.06). Again, this goes against the result we are looking for, since it should reduce the high subgroup's contributions in the u-shaped compared to the low-variance treatment.

common they thought the shown distribution was on a scale from 1 to 7. We can then split participants into two groups — those who think it is rather common (5–7) and those who think it is rather uncommon (1–4). Table B.2 shows that controlling for this variable does not affect the result on the importance of variance. What is more, whether participants perceive the distribution as common or not has no significant effect on contribution decisions.

Finally, our results also hold when we use average individual beliefs and the standard deviation of individual beliefs in Part II as regressors instead of the mean and variance of the observed distribution. This allows us to control for possible confounding effects arising from misperceptions of the observed distribution. Table B.3 shows that our results continue to apply.[14] In the high contribution sub-sample, a higher standard deviation in individual beliefs has a positive effect on contributions, while it has a negative effect in the low contribution sub-sample.

Table B.3: Tobit models. Effect of variance in beliefs on contributions in Part II for different sub-samples

	(1) High contributions (≥2)	(2) Low contributions (≤2)
Average individual belief in Part II	0.64***	0.36***
	(0.06)	(0.07)
SD of individual beliefs in Part II	0.09*	-0.10*
	(0.05)	(0.05)
Constant	0.45	-1.09**
	(0.46)	(0.47)
Baseline controls	Yes	Yes
Demographic controls	Yes	Yes
N observations	906	602
Pseudo R^2	0.16	0.08

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.
Note. Data is censored at 0 and 4. The dependent variable is *contributions in Part II.* Baseline controls include average individual beliefs, SD of individual 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 GPS 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). Regressions control for order effects

B.8 Personal traits and demographics

In the experiment, we measure personal values twice, 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". Table B.4 shows that our results hold also 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

[14] A hybrid approach where we control for individual average beliefs but insert high-variance and u-shaped dummies instead of the standard deviation of beliefs delivers an equivalent result.

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.4: Tobit models. Effect of personal traits on contributions in Part II (controlling for change in PVs)

	(1)	(2)
High mean	0.63***	0.66***
	(0.11)	(0.11)
Variance (baseline = low)		
High variance	-0.41	-0.53
	(0.50)	(0.50)
U-shaped	-0.86	-0.94*
	(0.54)	(0.53)
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-riding aversion x high variance	0.05	0.04
	(0.07)	(0.07)
Free-riding 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)
PVs x high variance	0.23*	0.26**
	(0.13)	(0.13)
PVs 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. Results of two tobit regressions (data is censored at 0 and 4). The dependent variable is *contributions in Part II*. *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. *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. *Change in personal values* is the difference between PVs in Part II and I and takes values between -4 and 4. Demographic controls include age, gender, education, acceptance of risk, trust, and GPS measures for negative and positive reciprocity. All regressions control for order effects.

Finally, we find that some demographic variables correlate with personal traits. In Table B.5 we regress personal values in column (1), sucker aversion in column (2), and free-riding aversion in column (3) on demographic characteristics. We find that women and younger participants have both a higher sucker and a higher free-riding aversion. Personal values on the other hand are slightly lower for female participants.

Table B.5: 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. Results of three tobit regressions. The dependent variable in column (1) is *personal values* in Part I and is censored at 0 and 4. The dependent variables in column (2) and (3) are *sucker* and *free-riding aversion* and are measured on a Likert scale from 1 to 7.

B.9 Contributions in Part I

As mentioned in the discussion section, unconditional contributions in Part I are an alternative measure for an individual's cooperative tendencies. Unsurprisingly, they are highly correlated with personal values. As a robustness check we can perform the same analysis as in Table 2 but using contributions in Part I instead of personal values as a measure for an individual's unconditional cooperative tendencies. As Table B.6 shows our main results are qualitatively unchanged. Both free-riding and sucker aversion matter more in the u-shaped than in the low-variance condition and just like personal values, contributions in Part I are more predictive of Part II behavior in the high-variance and the u-shaped than in the low-variance condition.

Table B.6: Tobit models. Effect of personal traits on contributions in Part II - using contributions in Part I as a measure for unconditional cooperative tendencies

	(1)	(2)
High mean	0.95***	0.96***
	(0.10)	(0.10)
Variance (baseline = low)		
High variance	-0.70	-0.71*
	(0.43)	(0.43)
U-shaped	-1.05**	-1.19***
	(0.46)	(0.45)
Sucker aversion	-0.14***	-0.13**
	(0.05)	(0.05)
Sucker aversion x high variance	0.01	0.04
	(0.07)	(0.07)
Sucker aversion x u-shaped	-0.18**	-0.17**
	(0.08)	(0.07)
Free-riding aversion	0.14***	0.14***
	(0.04)	(0.05)
Free-riding aversion x high variance	-0.00	-0.02
	(0.06)	(0.06)
Free-riding aversion x u-shaped	0.23***	0.23***
	(0.07)	(0.07)
Contributions in part I	0.61***	0.58***
	(0.07)	(0.07)
Contributions x high variance	0.34***	0.34***
	(0.10)	(0.10)
Contributions x u-shaped	0.61***	0.65***
	(0.11)	(0.11)
Constant	0.49	-0.77
	(0.32)	(0.60)
Demographic controls	No	Yes
N observations	1203	1188
Pseudo R^2	0.17	0.18

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

Note. Results of two tobit regressions (data is censored at 0 and 4). The dependent variable is *contributions in Part II*. *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. *Sucker* and *free-riding aversion* are measured on a Likert scale from 1 to 7. *Contributions in Part I* are used as an alternative measure for an individual's basic cooperative proclivities and take values between 0 and 4. Demographic controls include age, gender, education, acceptance of risk, trust, and GPS measures for negative and positive reciprocity. All regressions control for order effects.

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 event 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 first terms in (9). 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)

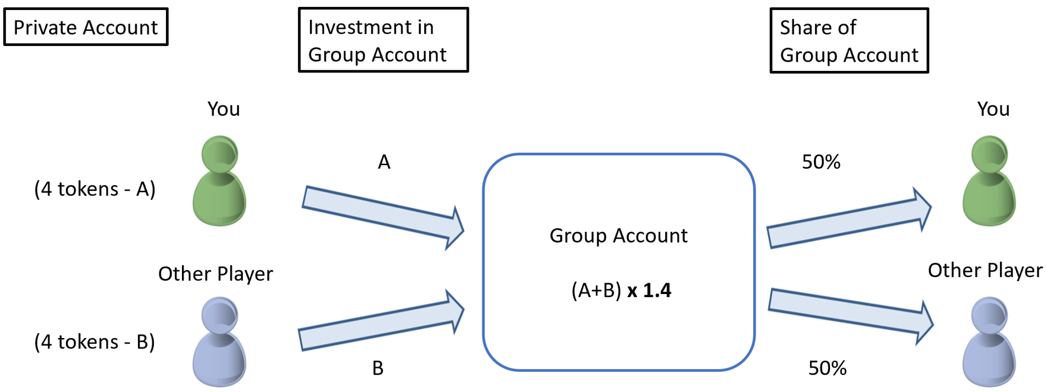
In this study, you will be anonymously paired with another participant. You will each start with **4 token** in your personal **private accounts**. In addition to the private accounts, there is a **group account**. You have to decide how many of your token you want to invest in the group account (either 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.

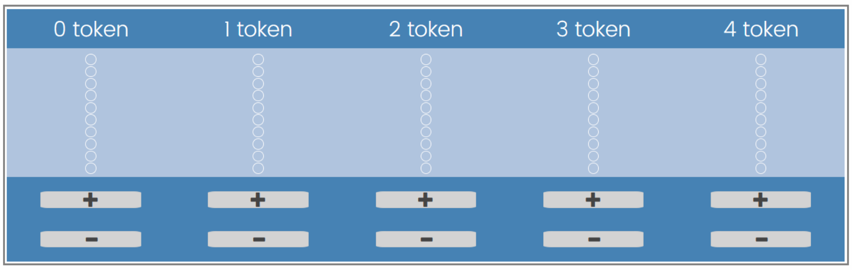
Your total income

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.[15] 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.

[15] Note that this is a common setup 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:

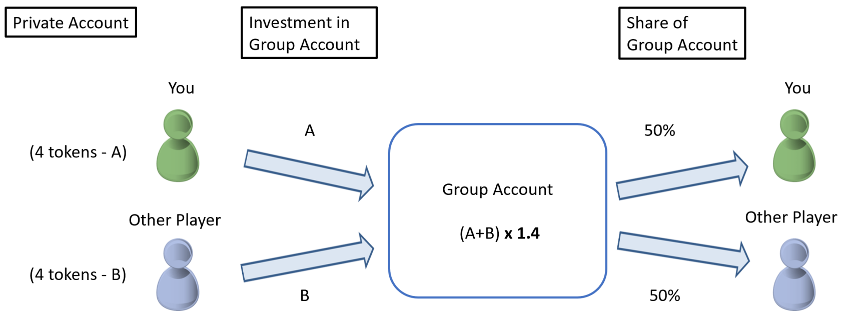
Your earnings are calculated on the basis of the table below. The more points you put on the correct option the higher your earnings. For each wrong option to which you allocate points your earnings will be reduced. The reduction is larger the more points you allocated to that option.

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.[16]

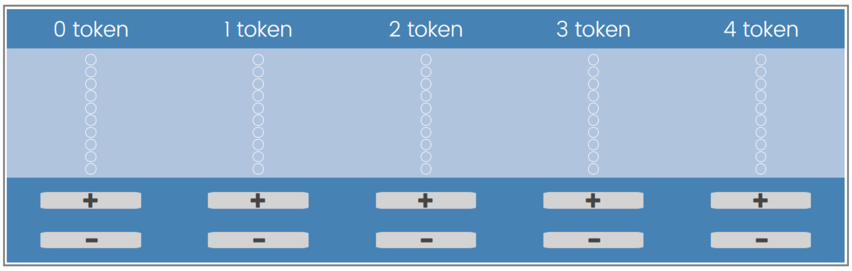
[16]Either normative questions or ABC questions are asked first. The three normative questions appear in randomized order.



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...

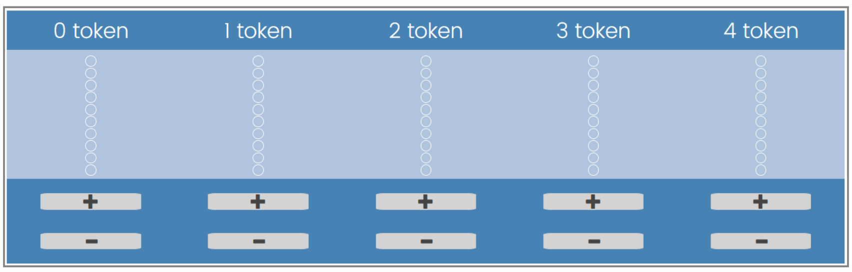


How confident are you in your response above? (0 not very confident, 100 very confident)

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.

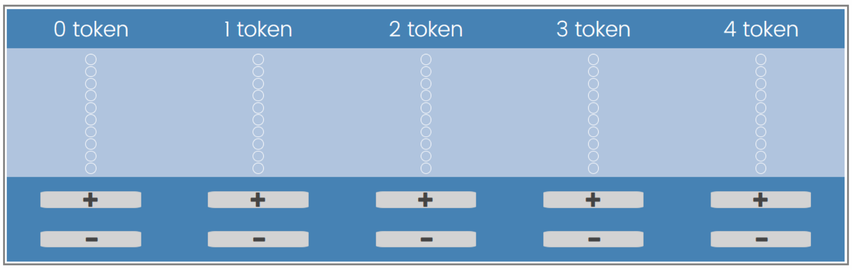
1) How many token do you want to **invest** in the **group account**?

- 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:

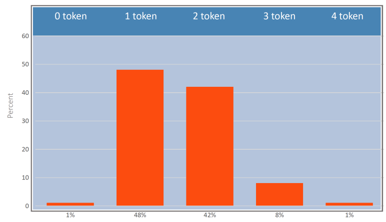
- The other player invests 0 token: ______ token
- The other player invests 1 token: ______ token
- The other player invests 2 token: ______ token
- The other player invests 3 token: ______ token
- The other player invests 4 token : ______ token

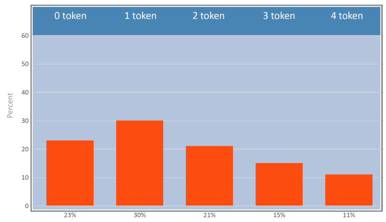
Part 2

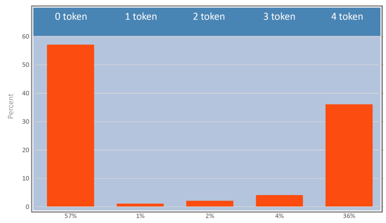
In a previous study we asked over 600 participants to **make an investment decision** in the same situation. The possible choices were to invest 0, 1, 2, 3 or 4 token in the **group account**. From their answers we constructed different subgroups. The graph below shows the percentage of people choosing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

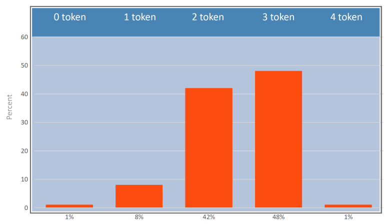
What previous participants invested in the group account:

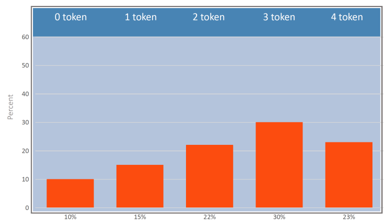
(Participants are randomly shown one of the following six pictures.)

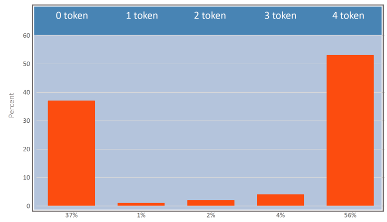












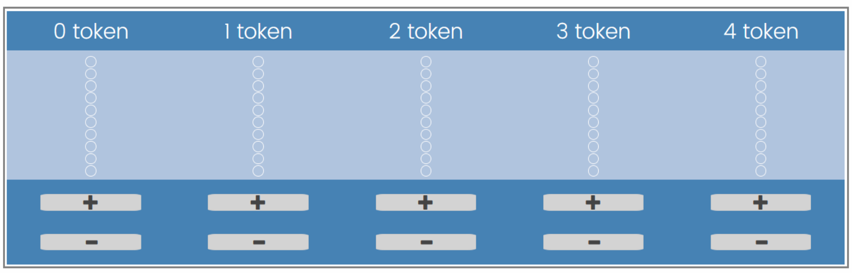
For Part 2 of the experiment you are matched with **one of the participants from the subgroup above**. You will only interact once with this person and you will never learn each other's identity.

Your and the other participant's **bonus payment for Part 2** will depend on your decisions and the decisions of this participant.[17]

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...



[17]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

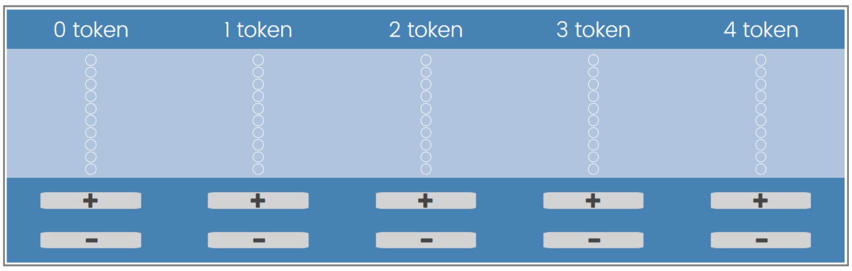
3) How many token do you want to **invest** in the **group account**?

- 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...

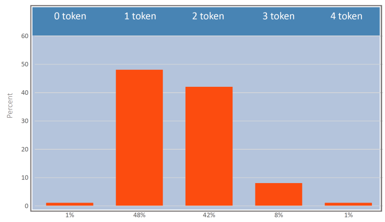


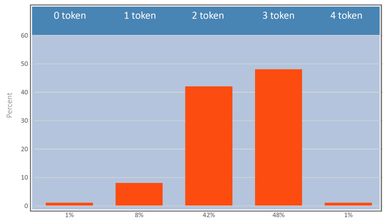
How confident are you in your response above? (0 not very confident, 100 very confident)

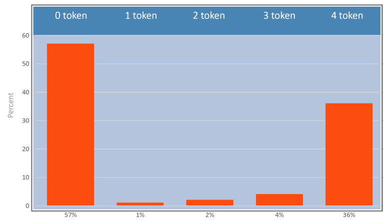
Questionnaire (1/2)

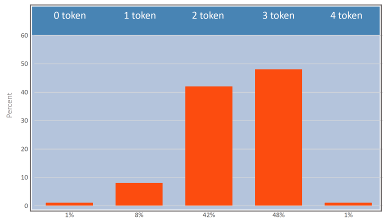
In this survey we showed you how many token a subgroup of other participants **invested** in the **group account**. Their answers are represented by the graph below.

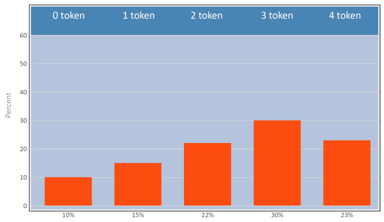
What previous participants invested in the group account: *(Participants are randomly shown one of the following six pictures.)*

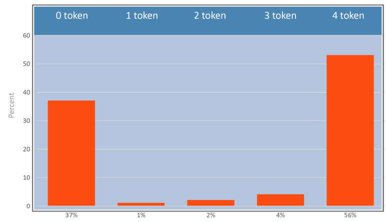












We will now ask you a few questions about the graph.

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

Taken from Haerpfer et al. (2022).

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) *Taken from Falk et al. (2023).*

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) *Taken from Falk et al. (2023).*

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) *Taken from Dohmen et al. (2011).*

5) What is your age?

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: