The Role of Fairness Ideals in Coordination Failure and Success

ANDRZEJ BARANSKI

New York University Abu Dhabi, e-mail: a.baranski@nyu.edu

ERNESTO REUBEN

New York University Abu Dhabi, Center for Behavioral Institutional Design, Luxembourg Institute of Socio-Economic Research, e-mail: ereuben@nyu.edu

ARNO RIEDL

Maastricht University, IZA Institute of Labor Economics, Center for Economic Studies and Ifo Institute, Netspar, e-mail: a.riedl@maastrichtuniversity.nl

ABSTRACT

In a laboratory experiment, we study the role of fairness ideals as focal points in coordination problems in homogeneous and heterogeneous groups. We elicit the normatively preferred behavior about how a subsequent coordination game should be played. In homogeneous groups, people share a unique fairness ideal for how to solve the coordination problem, whereas in heterogeneous groups, multiple conflicting fairness ideals prevail. In the coordination game, homogeneous groups are significantly more likely than their heterogeneous counterparts to sustain efficient coordination. The reason is that homogeneous groups coordinate on the unique fairness ideal, whereas heterogeneous groups disagree on the fairness ideal to be played. In both types of groups, equilibria consistent with fairness ideals are most stable. Hence, the difference in coordination success between homogeneous and heterogeneous groups occurs because of the normative disagreement in the latter types of group, making it much harder to reach an equilibrium at a fairness ideal.

This version: October 2025

ACKNOWLEDGEMENTS

We thank the audiences of multiple seminars and workshops where this paper was presented. We also acknowledge Mark Bernard's contribution in the early stages of the project. Baranski and Reuben gratefully acknowledge financial support from Tamkeen under the NYU Abu Dhabi Research Institute Award CG005.

1. Introduction

In the economy and society, pursuing a common goal often requires the coordination of actors who are heterogeneous in several aspects. Individuals in teams frequently have different abilities and talents, firms collaborating in R&D may have different development costs and receive different benefits from the created innovation, and cooperating nations commonly differ in wealth, population size, and may receive different benefits from the cooperation.

For example, in the North Atlantic Treaty Organization (NATO), some member countries have a limited capacity to contribute to the common defense goal but may derive a higher benefit than others from successfully coordinating on military funding levels. The recent fierce discussions about the minimal defense budget reflect the salience of the coordination problem. Similarly, in the European Union (EU), the creation of the European Stability Mechanism (ESM), in response to the Eurozone debt crisis between 2009 and 2010 (Schoenbaum, 2012), required funding to lend money to member states in need. During the creation of the ESM a heated debate between member states ensued about how to fund the mechanism, clearly emanating from the members' heterogeneity of perceived benefits and capacity to provide funding. In the same vein, in the challenge to mitigate climate change, a global debate has ensued about burden sharing associated with reparations, mitigation, and damages compensation related to environmental degradation, which is fueled by the heterogeneity of nations regarding their contribution to greenhouse gases, wealth, and proneness to negative and positive effects of climate change, among other things.

In this paper, we investigate coordination problems akin to those described above in homogeneous and heterogeneous groups. More specifically, with the help of a laboratory experiment, we examine the role of *fairness ideals* as focal points in coordination success and failure in both types of groups.[1]

It has been argued in the economic literature that focal points can be a powerful device to overcome coordination problems. Starting from the seminal work of Schelling (1960), it has been shown that focal points can effectively facilitate coordination, especially when the interests of the actors are aligned (see, e.g., Mehta et al., 1994; Crawford et al., 2008; Bardsley et al., 2010; Rojo-Arjona et al., 2022, among others), and that efficiency and fairness considerations can be successful facilitators of coordination (Isoni et al., 2014). However, in coordination problems

[1] Coordination problems have been widely investigated in the experimental literature, often using the weakest-link (minimum effort) games, where the lowest contributor determines the level of provision of the public good. Following the seminal work of Hirshleifer (1983) and Van Huyck et al. (1990), which showed that efficient coordination may be difficult to sustain, a large literature has emerged to understand how to achieve efficient coordination; see, for example, Weber (2006) for the role of group size, Brandts and Cooper (2006) and Hamman et al. (2007) for the role of incentives, Riedl et al. (2016) for partner choice, and Avoyan and Ramos (2023) for communication and gradual contributions. None of these studies examines the role of fairness ideals and normative (dis)agreement in homogeneous and heterogeneous groups.

like the ones described above, the role of efficiency and fairness as a coordination device may be limited because efficient outcomes are often not unique, and in heterogeneous groups, ideas about what constitutes a fair outcome may differ among actors. Indeed, in allocation problems, it has been found that people subscribe to a plurality of fairness ideals (Cappelen et al., 2007) and in public goods games with heterogeneous actors, different fairness norms are enforced (Reuben and Riedl, 2013). However, little is known about the existence and potential effects of a plurality of fairness ideals in coordination problems. We fill this gap and investigate whether there is a plurality of fairness ideals in a coordination problem with homogeneous and heterogeneous groups, respectively, leading to ex ante *normative disagreement*, and if the existence of one or more fairness ideals affects coordination success and failure in both types of groups.

We conduct a controlled and incentivized laboratory experiment and use a threshold public goods problem as our workhouse. In such a problem, the provision of the public good hinges upon surpassing a collective threshold with individual voluntary contributions. Contributions exceeding this threshold become redundant, adding no further value to the public good, while those falling short are irretrievably lost if provision fails.[2]

We use a threshold public goods problem for two distinct reasons. First, the threshold nature allows us to implement a severe coordination problem. As detailed further below, this public goods problem has one inefficient equilibrium where no one contributes, but a plethora of efficient equilibria, which makes the coordination problem particularly interesting as it discards efficiency as a possible focal point. Moreover, it allows us to create a situation where homogeneous and heterogeneous groups potentially differ in the plurality of fairness ideals. Second, the threshold characteristic resembles numerous real-world scenarios, like the ones outlined above, in which success depends on aggregating a critical mass of resources. For instance, take NATO's mission to deter military aggression, where the efficacy of military investments hinges on achieving a certain threshold: insufficient investments are wasteful as they render the alliance ineffective, while excessive investments may offer no additional deterrence benefits. Regarding the ESM, the treaty to fund it was informally called "Big Bazooka" implying that the funding capacity needed to be large enough to calm fears about debt default (i.e., a threshold for success must be met). Finally, in the climate change mitigation problem, the effectiveness of measures to reduce greenhouse gas emissions relies on widespread participation of nations and has been identified by scholars as a dilemma inherently having a threshold nature (Alley et al., 2003).

Specifically, we implement a 4-person threshold public goods problem, with a baseline treat-

[2]Threshold public goods experiments have a long history. The first ones by Hardin (1976) and Van de Kragt et al. (1983) consisted of binary, "all-or-nothing" contribution games. In this context, Rapoport (1988) explores endowment heterogeneity and suggests that differences in fairness views may drive differences in contributions. For experiments with a richer set of contribution choices, see Isaac et al. (1989), Bagnoli and McKee (1991), and Rapoport and Suleiman (1993). The latter two experiments introduce heterogeneity in benefits and endowments, respectively. For an experiment with communication, see Palfrey et al. (2017) and for a meta-analysis, see Croson and Marks (2000).

ment consisting of groups of players who are homogeneous with respect to the important controllable aspects of the experiment.[3] In homogeneous groups, all players have the same initial endowment and receive the same benefit from the public good when it is provided. In addition, there are two treatments with different types of heterogeneous groups. In one treatment, players differ in their initial endowment, with two having a low endowment and two having a high endowment, but all receive the same benefit from the provided public good. In the other treatment, all have the same endowment but differ in the benefit they receive from the public good: two receive a low benefit, and two receive a high benefit. All heterogeneity is public information, and the set of Nash equilibria is identical in the homogeneous and heterogeneous treatments.

As we aim to understand the role of fairness ideals for coordination, an important and innovative aspect of our experiment is the elicitation of those ideals. We do this in an incentive-compatible way using a variant of the spectator (benevolent dictator) method introduced by (Konow, 2000), which works as follows. Players are informed about the threshold public goods game, and they know both their endowments and benefits derived from the public good. Next, before they play the public goods game themselves, subjects are asked to implement their preferred contribution vector in another group (with which they will not interact). We refer to this choice as a subject's *normative contribution vector*, which arguably captures their fairness ideal. We also elicit players' beliefs about others' normatively preferred contribution vectors. Thereafter, subjects play the threshold public goods game repeatedly in the same group. These elicitations allow us to observe players' fairness ideals, their expected normative contribution vectors, and whether there is any normative disagreement. Finally, we examine how coordination failure and success differ between group types and, importantly, whether they are linked to fairness ideals.

Our main results are as follows. First, in homogeneous groups, we observe a unique fairness ideal that is consistent with the equality and efficiency norm: all players should contribute equally in a way that ensures the efficient provision of the public good. In stark contrast, in heterogeneous groups, we see a multiplicity of fairness ideals. Although there is agreement that the public good should be provided efficiently, there is considerable disagreement about the relative contributions of the different types of players. Interestingly, the most prominent normative contribution vectors are fairness ideals that appeal to the ideas of equality and proportionality, but focus on different aspects of the environment (e.g., contributions vs. earnings), which leads to normative disagreement. Second, homogeneous groups are more successful in providing the public good than heterogeneous groups. Interestingly, the main difference between these groups

[3] For reviews of the standard linear public goods game see (Ledyard, 1995; Zelmer, 2003; Chaudhuri, 2011) and Anderson et al. (2008), Buckley and Croson (2006), Fischbacher et al. (2014), and Reuben and Riedl (2013) for various forms of heterogeneity.

is the ability to sustain the public good once it is provided, which is significantly smaller in heterogeneous than in homogeneous groups. Third, fairness ideals play an important role in the coordination on public good provision. In homogeneous groups, almost all successful coordination is based on the unique fairness ideal, whereas in heterogeneous groups, coordination on one of the multiple fairness ideals is much less prevalent. Fourth, in all treatments, contributions at any fairness ideal are much more stable than (successful) contributions at other contribution vectors. Together with our third point, this explains why homogeneous groups are more successful in efficiently providing the public good.

Our study contains several novel aspects that contribute to a better understanding of coordination behavior. First, to the best of our knowledge, we are the first to elicit (expected) judgments about the normatively preferred way of behavior in coordination games with homogeneous and heterogeneous players and relate those to actual behavior in the coordination problem.[4] Second, combining the data from normative contributions and strategic interactions in the threshold public goods game allows us to assess whether actual choices in groups are indeed affected by individuals' own and expected normative contribution vectors, and whether ex-ante normative disagreement spills over into actual conflict. Third, our relatively long time horizon (the game is repeated for 20 periods) allows us to analyze the *stability* properties of different contribution vectors. Finally, we study different types of heterogeneity, which allows us to investigate whether fairness ideals and behavior vary depending on the source of heterogeneity.

With our work, we also contribute to the literature on focal points in coordination games[5] by showing that fairness ideals can act as such salient factors in the environment and contribute to both coordination success (in homogeneous groups) and coordination failure (in heterogeneous groups). We also contribute to the extensive and growing research agenda on fairness and equity principles, which has attracted the attention of social psychologists (Adams, 1965; Mellers and Baron, 1993), sociologists (Cook and Hegtvedt, 1983), and economists (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000; Konow, 2000, 2003; Cappelen et al., 2010, 2013; Cettolin and Riedl, 2017; Cappelen et al., 2022). This literature strongly focuses on individual decisions, whereas we explore the much less understood effect of the heterogeneity of fairness ideals in strategic settings like coordination problems.

The remainder of the paper is organized as follows. Section 2 describes the design and procedures, including the task for eliciting fairness ideals and the threshold public goods game. Section 3 discusses the equilibrium selection problem and introduces the fairness ideals. Sec-

[4]Bardsley et al. (2010) apply a related technique to two-player pure coordination games but do not elicit normatively preferred behavior or fairness ideals. Extending Konow (2000), Cappelen et al. (2007) and Cappelen et al. (2010) use a dictator game preceded by a joint production stage to simultaneously estimate the prevalence of three ideals of distributive justice among players and the distribution of the weight they attach to fairness.

[5]See, Schelling (1960) and subsequent theoretical and empirical studies, such as Sugden (1993, 1995), Mehta et al. (1994), Bacharach (1993), Bacharach and Bernasconi (1997), Bacharach and Stahl (2000), Crawford and Haller (1990), Janssen (2000), Blume and Gneezy (2000), and Isoni et al. (2014).

tion 4 presents the results, and Section 5 concludes.

2. Design and Procedures

In an experiment, we investigate how people solve coordination problems in homogeneous and heterogeneous groups and what role fairness ideals based on normative judgments play in this process. Our empirical strategy consists of two main components. First, we elicit in an incentive-compatible way subjects' normative judgments of how a coordination game should be solved, including the payoff consequences of their choices. Second, subjects interact repeatedly in the coordination game. In the analysis, we link the normative judgments with the actual choices. In the following, we describe these two design components in detail.

2.1. The threshold public goods game

In the experiment, our workhorse is a threshold public goods game, a prominent coordination game with many applications outside the laboratory. Subjects interact in the threshold public goods game in groups of four players. Each player i receives an endowment of y_i points. Players simultaneously decide how many points c_i to contribute to the provision of a public good, with $c_i \in \{0, 1, \ldots, y_i\}$. The public good is provided if the sum of contributions, $\sum_i c_i = C$, equals or surpasses a threshold Γ. If provided, player i receives a fixed benefit $b_i\Gamma$ points from the public good. Any contributions to the public good are sunk. In other words, individual i's earnings are given by

$$\pi_i = \begin{cases} y_i - c_i & \text{if } C < \Gamma, \\ y_i - c_i + b_i\Gamma & \text{if } C \geq \Gamma. \end{cases}$$

Note that if $\sum_j b_j \geq 1$, then providing the public good contributing such that $C = \Gamma$ is efficient. Moreover, if $b_i < 1\ \forall i$ or $y_i < \Gamma\ \forall i$, then nobody has the incentive or capacity to be the sole contributor to the public good. These conditions will hold in all of our treatments, as explained below.

2.2. Treatments

We implement three treatments. First, in the *Homogeneous* treatment (H for short), all group members have the same endowment, $y = 30$ points, and receive the same marginal benefit from the public good, $b = 0.5$.[6] In the treatment with *Heterogeneous Endowments* (HE), each group contains two high types with an endowment of $y^H = 60$ points and two low types with an endowment of $y^L = 30$ points. In HE, the marginal benefit from the public good is $b = 0.5$ for all, as in H. In the treatment with *Heterogeneous Benefits* (HB), each group contains two

[6]For the ease of exposition, we drop the index i.

Table 1. Treatment parameters and number of observations

Treatment	Player type	Parameters y_i	b_i	Γ	$b_i\Gamma$	Number of groups	subjects
H	low	30 points	0.50	60 points	30 points	15	60
HE	low	30 points	0.50	60 points	30 points	16	32
	high	60 points	0.50	60 points	30 points		32
HB	low	30 points	0.50	60 points	30 points	16	32
	high	30 points	1.00	60 points	60 points		32

high types with a marginal benefit from the public good of $b^H = 1$ and two low types with a marginal benefit from the public good of $b^L = 0.5$. In HB, each player's endowment equals $y = 30$ points, as in H. In all treatments, we set the provision threshold at $\Gamma = 60$. The treatments are summarized in Table 1.

2.3. Eliciting normative judgments and playing the coordination game

At the beginning of the experiment, subjects are informed that they have been randomly distributed over groups of four and read a detailed description of the threshold public goods game of the treatment to which they are randomly assigned.[7] In HE and HB, they also learn whether they have been randomly assigned to be a high or a low type. They also know that their group and type will remain the same throughout the experiment. Subjects are informed that the experiment consists of two independent parts and that in the second part, they will make contribution decisions in the described threshold public goods game. They also know that they will receive more specific instructions about the second part after completing the first part. The instructions clarified that their choices in the first part will not affect their earnings in the second part.

Eliciting normative judgments

In the first part of the experiment, we use an allocation task to elicit subjects' normative judgments regarding contributions to the threshold public goods game. To do so, we use a variation of the so-called spectator (a.k.a. benevolent dictator) task introduced by Konow (2000).[8] Specifically, each subject chooses a feasible vector of contributions, $\mathbf{c} = (c_1, c_2, c_3, c_4)$. At the end of the experiment, one vector per group is randomly selected and implemented in another group anonymously. Thus, subjects know that their decision can have real payoff consequences for

[7] The Appendix contains detailed procedures and a sample of the instructions.

[8] In the task, subjects make decisions that affect only the earnings of other subjects. The task is incentivized in that decisions have payoff consequences for these other subjects. Among others, Cappelen et al. (2007, 2010) and Cettolin and Riedl (2017) successfully use this technique to elicit subjects' fairness and justice ideals in different settings.

others. When choosing the contribution vector, we ask them to put themselves in the position of a “neutral uninvolved arbitrator” and indicate the “appropriate” contribution of each group member. Importantly, in HE and HB, subjects can distinguish the high and low types in the other group. Before submitting their final decision, subjects can try out different contribution vectors to see the payoff consequences of their choices. Henceforth, we refer to the vector of contributions elicited this way as a subject’s *normative contribution vector*. Subjects in the other group are not informed of the choices until the end of the experiment.

After choosing their normative contribution vector, we ask subjects to estimate the normative contribution vectors of the other subjects in their group. That is, each subject submitted three expected contribution vectors, one for each fellow group member. Subjects earn 100 points for each contribution vector that they correctly estimate. Given this incentive scheme, subjects should report their *modal* estimate for each fellow group member. Subjects are not informed of the accuracy of their guesses until the end of the experiment. We call these estimates the subjects’ *expected normative contribution vectors*. We elicit expected normative contribution vectors to examine whether subjects expect normative (dis)agreement. Given that we are interested in the consequences of normative judgments on *coordination*, we elicit normative contribution vectors *after* subjects learned their type in HE and HB.[9]

Playing the coordination game

The second part of the experiment consists of playing 20 periods of the threshold public goods game described above. As mentioned, subjects stay in the same group of four subjects (partners matching) and keep their type throughout the game. At the end of each period, subjects see the amount contributed by each group member, their payoffs, and whether the public good was provided or not. After this part ended, subjects answered a post-experiment questionnaire and then were paid in private.

We conducted the experiment in the CESS laboratory of New York University. We used standard experimental procedures, including random assignment of treatments to sessions, anonymity, neutrally worded instructions, and monetary incentives. In total, 188 subjects participated in the 45-minute experiment. Each subject participated in only one treatment. Total compensation varied between $7 and $24, with an average of $16.13.

[9] Krupka and Weber (2013) elicit normative expectations using coordination games. Their procedure is designed for cases where subjects expect general agreement on the appropriateness of a given action. Hence, their procedure is not suitable for capturing normative disagreement. Similar to our approach, Bicchieri and Xiao (2009) elicit subjects’ normative beliefs and their expectations of the normative beliefs of others separately. Unlike our method, their elicitation of normative beliefs is purely cheap talk. For a discussion of the elicitation of normative beliefs, see Erkut and Reuben (2019).

3. Theoretical Considerations

Here, we illustrate the severity of the coordination problem and propose that fairness ideals can serve as focal points and help to sharply reduce it. However, even if fairness ideals are focal, a coordination problem remains in heterogeneous groups because of normative disagreement. For simplicity, we focus on the stage game and assume that players are risk-neutral and maximize their own earnings.

3.1. Multiple equilibria

In all treatments, in the stage game there is one inefficient equilibrium where no one contributes and the public good is not provided. Importantly, however, there are many equilibria where the public good is provided. Two conditions have to be satisfied for such an equilibrium. First, the public good has to be provided efficiently (i.e., the sum of contributions equals the threshold or $C = \Gamma$). Otherwise, players who contribute a positive amount can save costs by reducing their contribution while ensuring that the public good is still provided. Second, the payoff of each player needs to be weakly larger than their endowment (i.e, $c_i \leq 30$ for all i). If not, it is profitable for a player to deviate to no contribution. It can be shown that in each of the three treatments, 19,872 pure strategy Nash equilibria exist, all of which are efficient except for the non-provision equilibrium. The set of equilibria is the same in each treatment because the provision threshold is the same.

3.2. Fairness ideals as focal points

We propose that contribution vectors based on normative fairness principles, or *fairness ideals*, can serve as focal points that may considerably reduce the complexity of the coordination problem.[10] In the following, we assume that all players agree that efficiency is a desirable property and restrict our attention to efficient equilibria. Based on that, we put forward some prominent principles for fairness ideals and derive the associated contribution vectors in the

[10]Three other factors could mitigate the coordination problem: payoff dominance, risk dominance, and other-regarding preferences. However, none of them sufficiently reduces the number of equilibria. First, payoff dominance is not an effective equilibrium selection criterion given that all but one equilibrium are efficient. Second, applying the notion of generalized risk dominance (Peski, 2010), we find that this criterion does not reduce the number of equilibria where the public good is provided (for the proof see Appendix B). Third, allowing for other-regarding preferences may reduce the number of equilibria because equilibria with very unequal earnings may not exist anymore. Indeed, with such preferences, the number of equilibria decreases, but not sufficiently to effectively mitigate the coordination problem. For example, in the framework of Fehr and Schmidt (1999), if we assume that all subjects have unreasonably large utility losses from disadvantageous and advantageous inequality ($\alpha = 2$ and $\beta = 0.6$), more than a thousand equilibria remain in each treatment (H: 2,886; HE: 2,736; HB: 1,332). Note that the meta-synthetic averages are $\alpha = 0.43$ for disadvantageous inequality and $\beta = 0.29$ for advantageous inequality (Nunnari and Pozzi, 2022), suggesting that with realistic values for α and β, the number of equilibria is larger.

three treatments of the threshold public goods game.

The simplest fairness ideal in our context is *Equal Contributions*, which prescribes that all group members contribute 15 points. This vector of contributions is obviously attractive for homogeneous groups. For heterogeneous groups, it can be justified by a libertarian view that argues that the random allocation of different endowments and benefits does not provide a basis for differential treatment. Thus, the libertarian principle tolerates the earnings inequality that results from equal contributions in HE and HB (Cappelen et al., 2007).

An alternative notion of fairness is based on the equality of outcomes rather than contributions. We refer to this fairness ideal as *Equal Earnings*. In H, to attain equal earnings while providing the public good, all group members need to contribute 15 points. In contrast, in both HE and HB, high types must contribute 30 points, whereas low types contribute 0 points. This fairness ideal can be motivated by the accountability principle in the theory of justice proposed by Konow (1996), which posits that differences in outcomes are justified only when they emanate from variables under the control of the individual. Given that subjects' types are assigned exogenously, there is no reason for differences in earnings to arise from the provision of the public good under the accountability principle.

Finally, a third fairness ideal is *Proportionality*, which can be established relative to the sacrifice made to produce the public good or the benefits derived from it. We apply both separately because proportionality relative to endowments matters only in HE and relative to benefits only in HB.[11] In our treatment with heterogeneous endowments, high types have twice the endowment of low types. Hence, *equal proportional sacrifice* requires high types to contribute twice as much as low types, implying contributions of 20 points for high types and 10 points for low types.[12]

In the treatment with heterogeneous benefits, the notion of *equal proportional benefit* entails that high types contribute twice as much as low types since they benefit twice as much when the public good is provided. Hence, again, we have high types contributing 20 points and low types contributing 10 points. We note that the proportionality trivially prescribes equal contributions in H.[13]

Table 2 summarizes the fairness ideals just discussed and their implied fair contribution vectors. In H, all fairness ideals select the same unique Nash equilibrium of equal contributions. If,

[11] It is noteworthy that in each heterogeneous treatment the contributions prescribed by the notion of proportionality correspond to two different principles of fair taxation (Mill, 1871): the ability-to-pay and benefits principles (for a discussion of these principles, see Musgrave and Musgrave, 1989).

[12] This fairness ideal is the fundamental philosophical basis of the ability-to-pay principle (Kendrick, 1939), which is often considered a key component in determining how to fund international alliances and organizations and is considered central to the perceived fairness of tax systems.

[13] This fairness ideal underpins the ability-to-pay principle of fair taxation, which can be traced back to Adam Smith: "The subjects of every state ought to contribute towards the support of the government, as nearly as possible, in proportion to their respective abilities; that is, in proportion to the revenue which they respectively enjoy under the protection of the state" (see Book V, Chapter 2 in Smith, 1937).

Table 2. Nash Equilibria of the threshold public goods game that coincide with fairness ideals

	Treatment		
	H	**HE**	**HB**
Equal Contributions	$c_i = 15\ \forall i$	$c_i = 15\ \forall i$	$c_i = 15\ \forall i$
Proportionality	$c_i = 15\ \forall i$	$c_H = 20$ $c_L = 10$	$c_H = 20$ $c_L = 10$
Equal Earnings	$c_i = 15\ \forall i$	$c_H = 30$ $c_L = 0$	$c_H = 30$ $c_L = 0$

as hypothesized, fairness ideals serve as focal points, the coordination problem in this treatment is relatively simple, and subjects should quickly choose to contribute efficiently and equally despite the many other Nash equilibria. In HE and HB, fairness ideals reduce the number of equilibria to three. Hence, if fairness ideals serve as focal points, the coordination problem is simplified considerably relative to the set of Nash equilibria. However, and quite importantly, a coordination problem remains if subjects disagree on the fairness ideals. Therefore, coordination may be more difficult in heterogeneous than homogeneous treatments if normative disagreement exists about the fair contribution vectors.

We investigate several questions. First, do subjects' normative contribution vectors match the abovementioned fairness ideals? Second, is there agreement on a single normative contribution vector across all treatments, or is there disagreement, especially in the heterogeneous ones? Third, how well is the coordination problem in the threshold public goods game solved when subjects actually play the game? Fourth, are there differences between treatments? Finally, do fairness ideals influence the success or failure in overcoming the coordination problem?

4. Results

This section consists of three parts. First, we analyze subjects' own and expected normative contribution vectors; second, we study group coordination in the threshold public goods game, and third, we link the fairness ideals to actual coordination behavior.

4.1. Normative judgments

Each subject submits one normative contribution vector and three expected normative contribution vectors (one for each other group member) before they interact with each other. Therefore, our independent unit of observation to test for differences in normative contribution vectors is the individual vector, and for differences in expected normative contribution vectors, it is the average of the expected vectors a subject submitted. In this section, we use non-parametric

Table 3. Subjects' own and expected normative judgments

Note: Fraction of normative contribution vectors in which the public good is provided ($C \geq \Gamma$), provided efficiently ($C = \Gamma$ and $b_i\Gamma \leq c_i \, \forall i$), and provided at a fairness ideal. Normative disagreement equals the mean Euclidean distance between the (expected) normative contribution vectors of each subject and the centroid of all normative contribution vectors in a treatment.

	Own			Expected		
	H	**HE**	**HB**	**H**	**HE**	**HB**
Fraction providing the public good	0.983	0.984	1.000	0.994	0.974	0.964
Fraction providing the public good efficiently	0.983	0.953	0.969	0.978	0.948	0.948
Fraction providing the public good as a fairness ideal	0.933	0.719	0.766	0.867	0.755	0.818
Normative disagreement	1.536	8.073	9.963	2.172	7.536	9.766

tests to evaluate the statistical significance of treatment differences. Throughout the paper, we report p-values of two-tailed tests.

Table 3 summarizes some properties of subjects' own and expected normative contribution vectors by treatment. The first row of the table shows that in all treatments there is a consensus that the public good should be provided, and from the second row we can see that subjects think that the provision should be efficient (i.e., $C = \Gamma$ and $b_i\Gamma \leq c_i \, \forall i$). The fraction of efficient normative contribution vectors exceeds 0.953, and for expected normative contribution vectors it exceeds 0.948.[14] Table 3 also reveals that a clear majority of (expected) normative contribution vectors is consistent with one of the fairness ideals presented in Table 1. A fraction of at least 0.719 of normative contribution vectors and of at least 0.755 of expected normative contribution vectors coincides with one of these fairness ideals. These findings constitute our first result.

Result 1. *In all treatments, when subjects make normative judgments regarding contributions to the public good, almost all prescribe and expect others to prescribe contributions that efficiently provide the public good, and an overwhelming majority of these normative contributions correspond to fairness ideals.*

Given the prevalence of fairness ideals, it is important to know if subjects agree on the same fairness ideals or if there is normative disagreement and how this differs across homogeneous and heterogeneous groups. A first indication for differences across group types is that overall the fraction of normative contribution vectors corresponding to a fairness ideal is somewhat lower in HE and HB than in H. Statistically, equality of proportions across the three treatments is rejected for normative contribution vectors ($p = 0.004$ with a Fisher's exact test) and expected normative contribution vectors ($p = 0.049$ with a Kruskal-Wallis test).

[14] We do not find significant pairwise differences between treatments in the fraction of normative contribution vectors in which the public good is either provided or provided efficiently ($p > 0.483$ with Fisher's exact tests for normative contribution vectors and $p > 0.270$ with Mann-Whitney U tests for expected normative contribution vectors).

For deeper insight, Figure 1 visualizes the distribution of subjects' normative contribution vectors (panel A) and expected normative contribution vectors (panel B) by treatment. The scatter plots for HE and HB display the mean (expected) contribution prescribed to high (low) types on the vertical (horizontal) axis. The size of each circle reflects the observed frequency of the corresponding (expected) normative contribution vector. In H, the vertical and horizontal axes correspond to the mean (expected) contribution prescribed to two random low types.[15]

It is evident from the figure that, despite the commonalities described in Result 1, there are stark differences between homogeneous and heterogeneous groups concerning both normative contribution vectors and expected normative contribution vectors. Focusing first on the former, Figure 1A shows that while in H the vast majority of normative contribution vectors correspond to the unique fairness ideal of Equal Contributions (EC), in HE and HB there is considerable normative disagreement as normative contribution vectors are distributed over the three fairness ideals as well as some other contribution vectors. To quantify the degree of normative disagreement within a treatment, we compute the Euclidean distance between each normative contribution vector and the centroid of all normative contribution vectors in the treatment. Specifically, we order the normative contribution vector of each subject i, $\mathbf{c}_i = (c_{i1}, c_{i2}, c_{i3}, c_{i4})$, such that $c_{i1} \leq c_{i2} \leq c_{i3} \leq c_{i4}$ and then calculate $d_i = \sqrt{\sum_k (c_{ik} - \bar{c}_k)^2}$, where c_{ik} is the k^{th} element of i's normative contribution vector and $\bar{c}_k$ is the mean of the k^{th} element among all subjects in the same treatment as i. A distance for each expected normative contribution vector is calculated analogously and then averaged per subject.

Table 3 (row 4) reports the results and shows that normative disagreement is much lower in H (1.536) than in HE (8.073) and HB (9.963). A Kruskal-Wallis test rejects the hypothesis that the observed normative disagreement in the three treatments comes from the same distribution ($p < 0.001$). In addition, pairwise Mann-Whitney U tests indicate significantly less normative disagreement in H than in HE and HB, and significantly less normative disagreement in HE compared to HB ($p < 0.025$). Similar findings are obtained for expected normative contribution vectors (Kruskal-Wallis test, $p < 0.001$).[16]

It is noteworthy that normative disagreement in HE and HB is practically unchanged if we drop normative contribution vectors that do not correspond to a fairness ideal (the mean distance between normative contribution vectors becomes 7.902 in HE and 10.762 in HB). Given that these allocations comprise around three-quarters of all normative contribution vectors, we can assert that normative disagreement over fairness ideals is the main contributor to the

[15] Not much information is lost by averaging contributions by type since the fraction of normative contribution vectors that prescribe the same contribution to subjects of the same type is close to one. For normative contribution vectors, this fraction equals 0.950 in H, 0.969 in HE, and 0.969 in HB. For expected normative contribution vectors, this fraction equals 0.878 in H, 0.938 in HE, and 0.932 in HB.

[16] Pairwise Mann-Whitney U tests indicate significantly less normative disagreement in H (2.172) than in HE (7.536) and HB (9.766), and significantly less normative disagreement in HE compared to HB for expected normative contribution vectors ($p < 0.042$).

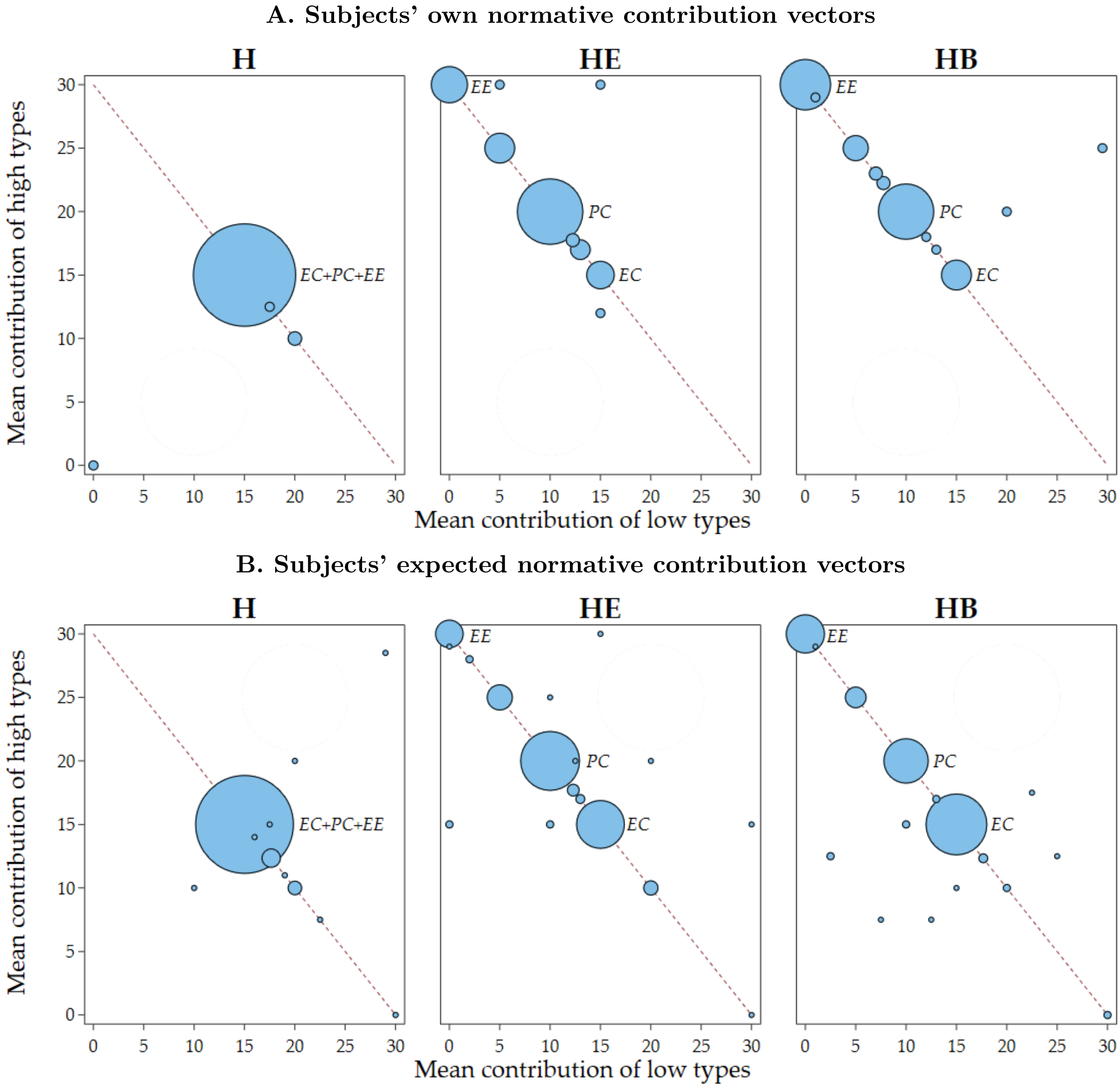

Figure 1. Subjects' own and expected normative contribution vectors

Note: The scatter plots depict subjects' own normative contribution vector (panel A) and expected normative contribution vectors (panel B). In HE and HB, the vertical (horizontal) axis corresponds to the mean (expected) prescribed contributions of high (low) types. In H, the vertical and horizontal axes correspond to the mean (expected) prescribed contributions of two random low types. Circles are drawn according to the frequency of each (expected) normative contribution vector. The dotted line corresponds to the set of efficient contribution vectors. Fairness ideals are indicated as EC for Equal Contributions, PC for Proportionality, and EE for Equal Earnings.

difference in normative disagreement between homogeneous and heterogeneous groups. These findings are summarized in our second result.

Result 2. *In homogeneous groups, there is overwhelming agreement on a unique normative contribution vector, whereas in heterogeneous groups, there is significant disagreement about what the normative contribution vector is.*

In the following, we check for two potential sources of normative disagreement in heterogeneous groups: self-serving bias and false consensus effect.

Table 4. Subjects' own and expected normative judgments by type

Note: Table includes only heterogeneous groups. Normative disagreement is the mean Euclidean distance between the (expected) normative contribution vectors of each subject and the centroid of all normative contribution vectors of *subjects of the same type*.

	HE		HB	
Subject's type	Low	High	Low	High
A. Share of total contributions prescribed to low types				
Own normative judgments	0.221	0.344	0.180	0.296
Expected normative judgments of other low types	0.271	0.325	0.268	0.335
Expected normative judgments of other high types	0.380	0.364	0.384	0.376
B. Normative disagreement				
Own normative judgments	8.315	6.108	9.942	8.094
Expected normative judgments of other low types	9.041	7.118	10.881	9.793
Expected normative judgments of other high types	8.007	6.429	8.963	9.731

In line with a number of other studies (see, e.g., Babcock et al., 1995; Gächter and Riedl, 2005; Karagözoğlu and Riedl, 2015), we find evidence that subjects make normative judgments in a self-serving manner. This can be seen in Table 4 (upper part), which shows—separately for low and high types—the share of the total contributions that subjects prescribe to the low types. On average, subjects assigned to be a low type think that low types should contribute 22.1% of the total contributions in HE and 18.0% in HB. By contrast, for subjects assigned to be a high type, the share prescribed to low types is 34.4% in HE and 29.6% in HB. This difference between high and low types is statistically significant in both treatments (Mann-Whitney U tests, $p < 0.012$). Interestingly, a similar pattern is observed for subjects' expected normative judgments. Specifically, low types expect that other low types will prescribe a significantly smaller share of total contributions to low types compared to the share they expect high types will prescribe (0.271 vs. 0.380 in HE and 0.268 vs. 0.384 in HB; Wilcoxon signed-rank tests, $p < 0.014$). The same effect, albeit smaller in magnitude, is seen for the expectations of high types (0.325 vs. 0.364 in HE and 0.335 vs. 0.376 in HB; Wilcoxon signed-rank tests, $p < 0.264$).[17]

Yet, even though self-serving biases are present, we find that disagreement *within* types is the more important source of normative disagreement. This can be observed when looking at normative disagreements within types. Comparing the normative disagreement when types are pooled (Table 3, row 4) with the normative disagreement separately for the types (Table 4, lower part, first row) reveals that the latter are not substantially different from the former for both types. Notably, the degree of disagreement within types in HE and HB remains substantially higher than in H (see Table 3, row 4; $p < 0.001$ with Kruskal-Wallis tests comparing the distribution of distances in H to the distribution of distances within each type in HE and HB).

[17] Although low and high types exhibit self-serving biases, Table C1 in the Appendix shows that they overwhelmingly agree that the public good should be provided efficiently and based on a fairness ideal.

The same pattern can be observed for expected normative contribution vectors. The degree of normative disagreement among high and low types when guessing the others' normative contribution vectors (Table 4, lower part, second and third row) is not substantially lower than the overall degree of disagreement in heterogeneous groups, and it is significantly higher than in homogeneous groups (see Table 3, row 4; Kruskal-Wallis tests, $p < 0.001$).

Next, we analyze whether the observed heterogeneity in expected normative contribution vectors in HE and HB is merely a false consensus effect (i.e., subjects believing others share their own normative judgments). To answer this question, we look at the fraction of subjects who expect that their own normative contribution vector matches the contribution vector of the other subject of the same type. This fraction is 56.3% in HE and 54.7% in HB. Hence, about half the subjects in both treatments expect some normative disagreement, even within types.[18] We summarize in our next result.

Result 3. *In heterogeneous groups, self-serving biases and a false consensus effect explain some of the normative disagreement between high and low types. However, most normative disagreement can be attributed to a lack of agreement within types over multiple fairness ideals.*

4.2. Coordination and public good provision

Table 5 reports descriptive statistics of the (efficiency of) groups' contribution vectors (i.e., the four contribution choices of subjects in a group in a period). It provides statistics separately for the first period, the first ten periods, and the last ten periods. Throughout this section, we use nonparametric tests to compare behavior across treatments. Our units of observation are group averages over all periods considered in a particular test.

We start by looking at the overall coordination success of homogeneous and heterogeneous groups. Since profits are difficult to compare across treatments due to differences in endowments and benefits from the public good, we use normalized efficiency gains to measure overall success. Specifically, we normalize each group's total profit per period such that zero equals the total profit when nobody contributes (i.e., the sum of endowments), and one equals the total profit when the public good is provided efficiently (i.e., the maximum group profit). Efficiency gains can be negative if subjects contribute positive amounts but fail to provide the public good. The first row of Table 5 reports the mean efficiency gain per group type. It shows that heterogeneous groups have considerably lower efficiency gains than homogeneous groups. In the first period, homogeneous groups already achieve 47.2% of the potential efficiency gains while heterogeneous groups achieve less than 11.4%. Over time, efficiency improves in all treatments, but the differences between H, on the one hand, and HE and HB, on the other hand, remain. Over

[18] There is evidence of false consensus as subjects overestimate the popularity of their own normative judgments. The mean probability of a subject being matched with someone who chose the same normative contribution vector is 25.3% in HE and 22.1% in HB.

Table 5. Efficiency of the groups' contribution vectors

Note: 'Mean efficiency gain' is the mean across periods of $\left(I_\Gamma \, \Gamma \sum_i b_i - C\right) / \left(\Gamma \sum_i b_i - \Gamma\right)$, where I_Γ is an indicator function that equals one if $C \geq \Gamma$. 'Fraction of provisions' and 'Fraction of efficient provisions' is the fraction of group contribution vectors in which the public good is provided ($C \geq \Gamma$) and provided efficiently ($C = \Gamma$ and $b_i \Gamma \leq c_i \, \forall i$), respectively.

	Period 1			Periods 1-10			Periods 11-20		
	H	**HE**	**HB**	**H**	**HE**	**HB**	**H**	**HE**	**HB**
Mean efficiency gain	0.472	-0.278	0.114	0.576	0.166	0.256	0.712	0.330	0.483
Fraction of provisions	0.733	0.313	0.375	0.747	0.525	0.450	0.813	0.581	0.619
Fraction of efficient provisions	0.467	0.125	0.063	0.653	0.281	0.144	0.760	0.400	0.419

the first ten periods, the treatment differences in efficiency gains are large and statistically significant (Mann-Whitney U tests, $p = 0.009$ for H vs. HE and $p = 0.037$ for H vs. HB), and these differences remain significant also in the last ten periods (Mann-Whitney U tests, $p = 0.046$ for H vs. HE and $p = 0.038$ for H vs. HB).[19]

The difference in efficiency gains between homogeneous and heterogeneous groups is due to both less frequent public good provisions and less efficient provisions in heterogeneous groups. These two effects can be observed in the second and third rows of Table 5. In the first period, compared to H, the fraction of provisions is at least 35 percentage points lower in HE and HB, and at least 34 percentage points lower if we look at the fraction of provisions that are also efficient. These differences remain large also in later periods. The differences in provision and efficient provision are statistically significant in the first ten periods (Mann-Whitney U tests, $p < 0.014$ for H vs. HE and $p < 0.013$ for H vs. HB). In the last ten periods, differences between homogeneous and heterogeneous groups narrow but remain statistically significant between H and HB (Mann-Whitney U tests, $p < 0.037$) and between H and HE for efficient provisions (Mann-Whitney U tests, $p = 0.033$; for provisions, the difference is significant at $p = 0.069$).[20]

To obtain additional insights concerning the persistent differences between homogeneous and heterogeneous groups in public good provision, we look at the provision dynamics. Specifically, we ask whether, compared to homogeneous groups, heterogeneous groups have more difficulties in reaching contribution vectors that provide the public good after having failed to do so, or sustaining contribution vectors that provide the public good after having succeeded in doing so, or both. The answer to this question is provided in Table 6. The table shows the fraction of groups that provide the public good in period $t + 1$, depending on whether they succeeded or failed in doing so in period t. We find that groups that *failed* to provide the public good in

[19] Differences between HE and HB in mean efficiency gains are not statistically significant in either the first or the last ten periods (Mann-Whitney U tests, $p > 0.439$).

[20] There are no statistically significant differences in the fraction of provisions between HE and HB in either the first or the last ten periods (Mann-Whitney U tests, $p > 0.526$). Compared to HB, the fraction of efficient provisions in HE is significantly higher at $p = 0.052$ in the first ten periods but not in the last ten periods ($p = 0.542$, Mann-Whitney U tests).

Table 6. Successful public good provision in period $t+1$ depending on provision in period t

Note: Fraction of groups that provide the public good in period $t+1$ ($C_{t+1} \geq \Gamma$) depending on whether they succeeded ($C_t \geq \Gamma$) or failed ($C_t < \Gamma$) to provide it in period t.

	H	**HE**	**HB**
Failed provision in period t	0.246	0.309	0.317
Successful provision in period t	0.929	0.774	0.748

period t are similarly likely to provide the public good in period $t+1$ in both homogeneous and heterogeneous groups (24.6% in H, 30.9% in HE, and 31.7% in HB; pairwise Mann-Whitney U tests, $p > 0.329$). By contrast, we see clear treatment differences in the fraction of groups that provide the public good in period t and keep providing it in period $t+1$. In homogeneous groups, the likelihood that a group keeps providing the public good is 92.9%, whereas, in heterogeneous groups, it is less than 77.4% (pairwise Mann-Whitney U tests, $p = 0.020$ for H vs. HE and $p = 0.011$ for H vs. HB). Interestingly, these differences are not due to differences in the fraction of failed provisions in the first period since groups in H that provided the public good in the first period have almost identical dynamics to groups that failed to provide it.[21] Hence, it appears that in comparison to homogeneous groups, the main problem for heterogeneous groups is not to attain a contribution vector at which the public good is provided. Instead, the problem is to sustain the provision of the public good.[22] These findings establish our fourth result.

Result 4. *Overall, heterogeneous groups are less successful than homogeneous groups in public good provision. Specifically, they are less likely to provide the public good and, when they do, they are less likely to do so efficiently. Both homogeneous and heterogeneous groups are equally adept at overcoming failure to provide the public good, but heterogeneous groups are less capable of sustaining public good provision.*

4.3. Fairness ideals and public good provision

We now turn to the investigation of the role of fairness ideals in explaining the differences in coordination success and failure between homogeneous and heterogeneous groups. We note first that in homogeneous groups, a unique fairness ideal is a prominent contribution vector. Already in the first period, groups in H coordinate on equal contributions 40.0% of the time, which accounts for 54.5% of all provisions. Over time, this unique fairness ideal becomes even more prominent: in the first (last) ten periods, it accounts for 60.0% (74.7%) of all outcomes

[21] Groups in H that provide the public good in period t are similarly likely to provide the public good in period $t+1$ irrespective of whether they provided the public good in the first period or not (93.2% vs. 91.7%). Analogously, groups in H that failed to provide the public good in period t are similarly likely to provide it in period $t+1$ irrespective of their first-period provision (24.2% vs. 25.0%).

[22] Note that if HE and HB had the same likelihood of staying successful as H (92.9%), then the initial differences in provision rates would have disappeared by period 6.

and 80.4% (91.8%) of all provisions. By comparison, in HE and HB, groups coordinate on a fairness ideal considerably less often. In the first period, groups in HE and HB coordinate on one of three fairness ideals less than 6.4% of the time, which accounts for less than 16.9% of all provisions. Over time, fairness ideals become more common, but they are never as dominant as in homogeneous groups: in the first (last) ten periods, fairness ideals account for less than 16.9% (24.4%) of all outcomes and 32.1% (41.9%) of all provisions. Pairwise Mann-Whitney U tests confirm that fairness ideals are played significantly less often in HE and HB than in H ($p < 0.002$ for the first ten periods and $p < 0.008$ for the last ten).[23]

Result 5. *In homogeneous groups, coordination behavior is mostly based on the unique fairness ideal of equal contributions. In heterogeneous groups, coordination on one of the three fairness ideals is significantly less prevalent.*

Next, we investigate whether the lower frequency of coordination on fairness ideals explains the lower ability of heterogeneous groups to sustain the provision of the public good. To answer this question, we look at the transition dynamics between intuitive categories of contribution vectors. Specifically, we classify contribution vectors into four categories: *failed* contribution vectors that do not provide the public good, *overprovided* contribution vectors that provide the public good inefficiently, *efficient* contribution vectors that provide the public good efficiently but are not a fairness ideal, and contribution vectors that provide the public good and are a *fairness ideal*. Then, we examine how groups transition between these categories by calculating the fraction of groups that played a contribution vector in a particular category in period $t+1$, depending on the category they played in period t.

These fractions are shown in Table 7 for both homogeneous and heterogeneous groups.[24] For example, the table shows that of all the homogeneous groups that failed to provide the public good in period t, in period $t+1$, 75.4% fail again, 9.8% manage to provide the public good inefficiently, 3.3% provide the public good efficiently but not at a fairness ideal, and 11.5% provide it at a fairness ideal.

To test whether differences between these fractions are statistically significant, we use a multinomial logit model with the category played in period $t+1$ as the dependent variable. As independent variables, we use dummy variables to indicate the category played in period t interacted with a dummy variable indicating whether a group is homogeneous or heterogeneous. We cluster standard errors on groups and use the estimation results to perform the tests reported below. Table C2 in the Appendix contains the regression estimates.

[23] Importantly, the low frequency of fairness ideals is not due to subjects consistently coordinating on other specific contribution vectors. In both types of heterogeneous groups, the modal contribution vector is still a fairness ideal: proportional contributions in HE and equal contributions in HB.

[24] We pool HE and HB for clarity of exposition. The fractions of these two treatments are nearly identical.

Table 7. Transition probabilities between different contribution vectors

Note: Contribution vectors are classified as either *failed* ($C < \Gamma$), *overprovided* ($C > \Gamma$), *efficient* ($C = \Gamma$ and $b_i\Gamma \leq c_i \,\forall i$, excluding fairness ideals), or *fairness ideal.* Each cell contains the fraction of groups that coordinated on a contribution vector in one of the categories in period t (rows) that subsequently coordinate on a contribution vector in one of the categories in period $t + 1$ (columns).

Panel A. Homogeneous

		Period $t+1$			
		Failed	Overprovided	Efficient	Fairness ideal
Period t	Failed	0.754	0.098	0.033	0.115
	Overprovided	0.227	0.227	0.091	0.455
	Efficient	0.300	0.200	0.500	0.000
	Fairness ideal	0.042	0.026	0.000	0.932

Panel B. Heterogeneous

		Period $t+1$			
		Failed	Overprovided	Efficient	Fairness ideal
Period t	Failed	0.687	0.213	0.068	0.032
	Overprovided	0.363	0.459	0.130	0.048
	Efficient	0.243	0.176	0.581	0.000
	Fairness ideal	0.065	0.028	0.000	0.907

The first notable finding is that *fairness ideal* is the most stable category in all types of groups. In both homogeneous and heterogeneous groups, more than 90.7% of the groups coordinating on *fairness ideal* in period t remain in *fairness ideal* in period $t+1$. Public good provision that does not occur at a fairness ideal is much less stable. Of those groups that had provided the public good in t, but not at a fairness ideal, at most 58.1% provide in the same category in $t + 1$. The differences in frequencies of *fairness ideal* with the other categories are significant (Wald tests, $p < 0.025$ for homogeneous groups and $p < 0.002$ for heterogeneous groups). Notably, among groups that are providing the public good in period t, groups that coordinated on *fairness ideal* are also less likely to switch to *failed* in period $t + 1$ compared to groups that coordinated on *overprovided* or *efficient* (at most 6.5% in *fairness ideal* vs. at least 22.7% in the other two provision categories).[25] The differences between the *fairness ideal* and *efficient* categories are particularly interesting since subjects in these categories do not have a pecuniary incentive to change their contribution. In other words, provisions at a fairness ideal are much more stable than other (efficient) contribution vectors and this holds for both types of groups.

However, there is also an important difference between the two types of groups. Compared to homogeneous groups, heterogeneous groups are much less likely to transition from the *over-*

[25] Among heterogeneous groups, the fraction switching to *failed* is significantly lower for *fairness ideal* compared to both *overprovided* and *efficient* (Wald tests, $p < 0.031$). In homogeneous groups, this fraction is significantly lower compared to *efficient* but not compared to *overprovided* (Wald tests, $p < 0.002$ and $p = 0.146$).

provided to the *fairness ideal* categories (45.5% vs. 4.8%, $p = 0.004$).[26] In other words, it is much harder for heterogeneous groups to reach one of the fairness ideals.

In our next result, we summarize the two main findings explaining the lower ability of heterogeneous groups to sustain (efficient) coordination.

Result 6. *Once reached, fairness ideals are significantly more stable than other contribution vectors in both homogeneous and heterogeneous groups. However, coordination on fairness ideals is harder to attain in heterogeneous groups.*

4.4. Individual behavior and fairness ideals

In this section, we analyze contributions at the individual level to better understand the role that fairness ideals and normative judgments play in the choice and stability of contribution vectors. We also examine how group-level heterogeneity impacts public good provision and its stability.

We start by looking at the fraction of individual contributions that are consistent with (i) a fairness ideal, (ii) the subjects' own normative judgment, and (iii) their expected normative judgments. We categorize a contribution as consistent with a subject's *fairness ideal* if it equals 15 points in homogeneous groups, 0, 10, or 15 points for low types in heterogeneous groups, and 15, 20, or 30 points for high types in heterogeneous groups. Similarly, we categorize a contribution as consistent with a subject's *own normative judgment* if it equals the contribution prescribed by that subject to players of her type (e.g., if i is a low type and her normative contribution vector is proportional contributions, then a contribution of 10 points is consistent with i's normative judgment).[27]

We categorize consistency with *expected normative judgments* similarly. However, as subjects submit three expected normative contribution vectors (one for each other group member), instead of being categorized as either consistent or inconsistent, a contribution is categorized as the fraction of the subject's expected normative judgments it is consistent with.

Table 8 reports the outcome of this analysis for period 1, periods 1-10, and periods 11-20. For period 1, the table reveals two notable observations about contributions. First, subjects in heterogeneous groups choose a contribution consistent with a fairness ideal as often as subjects in homogeneous groups (pairwise Fisher's exact tests, $p = 0.299$ for H vs. HE and $p = 0.474$ for

[26] Wald tests comparing homogeneous and heterogeneous groups across each cell in Table 7 result in few statistically significant differences. Other than the difference highlighted in the text, there are two other statistically significant differences. Namely, homogeneous groups are less likely than heterogeneous groups to transition from *failed* to *overprovided* (9.8% vs. 21.3%, $p = 0.023$) and to remain in *overprovided* (22.7% vs. 45.9%, $p = 0.029$). All other differences are not statistically significant ($p > 0.176$).

[27] If i is one of the few subjects who prescribes different contributions to the two players of her own type (see footnote 15), then we categorize i's contribution as being not consistent with her normative judgment. Our results are unaffected if these subjects are dropped from the analysis.

Table 8. Characteristics of individual contributions

Note: Fraction of individual contributions that are consistent with: (i) a fairness ideal; (ii) the own normative judgment (i.e., the contribution prescribed to players of their type); (iii) the expected normative judgment (i.e., the contributions subjects expect others prescribe to players of their type).

	Period 1			Periods 1-10			Periods 11-20		
Fraction of contributions consistent with	**H**	**HE**	**HB**	**H**	**HE**	**HB**	**H**	**HE**	**HB**
... a fairness ideal	0.800	0.703	0.859	0.810	0.633	0.636	0.825	0.630	0.622
... the own normative judgment	0.800	0.547	0.359	0.787	0.327	0.261	0.795	0.298	0.255
... the expected normative judgments	0.753	0.424	0.430	0.754	0.293	0.260	0.751	0.240	0.277

H vs. HB). This suggests that the much lower fraction of groups coordinating on fairness ideals in heterogeneous groups relative to homogeneous groups (see Result 5) is not because subjects do not choose them initially, but because they choose different fairness ideals. This is illustrated by the fact that in period 1, only in 3.1% of the cases all group members contribute according to the *same* fairness ideal. Second, although an overwhelming majority chooses a fairness ideal, the fraction of subjects choosing according to their own normative judgments or expected normative judgments is significantly smaller in heterogeneous groups than in homogeneous groups (pairwise Fisher's exact tests for normative judgments, $p < 0.005$; pairwise Mann-Whitney U tests for expected normative judgments, $p < 0.001$).[28] Again, this is explained by subjects in homogeneous groups agreeing on a unique fairness ideal and acting according to it, while subjects in heterogeneous groups disagree and expect disagreement about which fairness ideal is the most acceptable.

Table 8 also shows that, after the first period, subjects in homogeneous groups continue to contribute according to the fairness ideals and their (expected) normative judgments to a large extent. In stark contrast, subjects in heterogeneous groups reduce the fraction of times they contribute according to fairness ideals. At the same time, although increasing, the number of cases that members in a group contribute according to the same fairness ideal stays relatively low (periods 1-10: 13.4%, periods 11-20: 22.2%). In addition, there is an even more noticeable decrease in the fraction of times subjects in heterogeneous groups contribute according to their own and expected normative judgments.

Thus, overall, in homogeneous groups subjects contribute according to the unique fairness ideal from period 1 onward. In heterogeneous groups, at the beginning subjects also contribute according to (multiple) fairness ideals. However, they fail to coordinate on the same fairness ideal and subsequently also contribute less often in accordance with fairness ideals, which leads to lower coordination success.

[28]This finding also holds if we concentrate solely on subjects whose normative contribution vector is a fairness ideal. Among these subjects, the fraction contributing in the first period according to a fairness ideal is 80.4% in HE and 89.8% in HB. However, the fraction contributing according to their normative judgment (expected normative judgments) is only 54.3% (41.3%) in HE and 46.9% (47.6%) in HB.

Table 9. Stability of individual contributions (probability that $c_{i,t} = c_{i,t+1}$)

Note: Logit regressions with subject random effects. The dependent variable equals one if $c_{i,t} = c_{i,t+1}$ and zero otherwise. Contribution vectors are classified as either *Failed* ($C_t < \Gamma$), *Overprovided* ($C_t > \Gamma$), *Efficient* ($C_t = \Gamma$ and $b_i\Gamma \leq c_{i,t}\,\forall i$, excluding fairness ideals), or *Fairness ideal.* All regressions contain 2432 observations from 128 subjects in 32 groups. The table reports marginal effects and robust standard errors clustered at the group level in parentheses. ** and * denote statistical significance at 1% and 5%, respectively.

	I	II
Failed × $c_{i,t}$ is a fairness ideal		0.094**
		(0.034)
Overprovided	0.019	−0.001
	(0.028)	(0.051)
Overprovided × $c_{i,t}$ is a fairness ideal		0.125*
		(0.050)
Efficient	0.242**	0.280**
	(0.058)	(0.100)
Efficient × $c_{i,t}$ is a fairness ideal		0.030
		(0.095)
Fairness ideal	0.446**	0.501**
	(0.033)	(0.043)
Constant	0.520**	0.466**
	(0.029)	(0.038)
Pseudo-R^2	0.150	0.092

One reason we have identified why heterogeneous groups are less successful than homogeneous groups in providing the public good is that they are less likely to sustain public good provision (see Result 4). This raises the question about the determinants of the stability of provision in these types of groups. To answer this, we investigate the relationship between the stability of individual contributions and group outcomes, on the one hand, and fairness ideals, on the other hand. In the following, we only look at heterogeneous groups because these relationships are more interesting when there is heterogeneity in fairness ideals. However, the findings reported below are not qualitatively affected when including homogeneous groups.

To study the stability of individual contributions, we look at subjects' decisions to keep contributing the same amount from one period to the next. We model this decision as a function of the group's contribution vector and whether the subjects' contribution is consistent with a fairness ideal. We estimate logit regressions where the dependent variable equals one if subject i contributes the same amount in period t and period $t+1$ (i.e., if $c_{i,t} = c_{i,t+1}$) and zero otherwise. In all regressions, we use subject random effects and cluster standard errors at the group level. The estimated marginal effects are presented in Table 9.

In regression I, we include dummy variables for the four categories of contribution vectors defined in Table 7 as independent variables. These variables capture, whether i's group in period t (i) *Failed* to provide the public good (which is the omitted category), (ii) *Overprovided* the public good, (iii) provided the public good in an *Efficient* way but not at a fairness ideal, or (iv) provided the public good at a *Fairness ideal.* Consistent with Result 6, subjects in the *Efficient* or *Fairness ideal* categories are significantly more likely to keep contributing the same amount than subjects in the *Failed* or *Overprovided* categories (Wald tests, $p < 0.001$). Moreover, subjects in the *Fairness ideal* category are significantly more likely to keep their contribution unchanged than subjects in the *Efficient* category (Wald test, $p = 0.004$).[29]

In regression II, we examine whether contributions consistent with fairness ideals are more stable even if groups do not coordinate on one. We use a dummy variable that equals one if i's contribution in period t was consistent with one of the fairness ideals (i.e., a low type who contributes 0, 10, or 15 points, or a high type who contributes 15, 20, or 30 points) and interact this dummy with the *Failed*, *Overprovided*, and *Efficient* categories. First, we observe that the effects of *Efficient* and *Fairness ideal* are robust to adding these interactions. Second, we find that subjects in groups that played a *Failed* or *Overprovided* contribution vector are between 9.8 and 12.5 percentage points more likely to keep their contribution unchanged if they contributed according to a fairness ideal ($p < 0.010$). Thus, contributing at a fairness ideal reduces the likelihood that a subject moves away from an unsuccessful or inefficient contribution vector.[30]

We summarize the main findings of the analysis in this section in the following result.

Result 7. *Initially, in both heterogeneous and homogeneous groups, individual contributions are equally likely to be at fairness ideals. However, in heterogeneous groups, there is disagreement about which fairness ideal is most appropriate, resulting in decreasing contributions at fairness ideals. Over time, individual contributions are most stable when groups coordinate on a fairness ideal. In addition, individual contributions are also more stable when they are consistent with a fairness ideal, even when groups overprovide or fail to provide the public good.*

This result highlights the key role fairness ideals play in both the success and failure of public good provision in homogeneous and heterogeneous groups. First, to achieve and sustain coordination, it is paramount to contribute at a fairness ideal. In homogeneous groups, this is relatively easy as there is normative agreement on a contribution vector. In heterogeneous groups, however, the normative disagreement over fairness ideals makes it more difficult to agree

[29] The results reported in this paragraph and Table 9 are without control variables. Adding controls for trends by including period fixed effects and time-invariant effects of specific contribution amounts by including a set of dummy variables $\alpha_x \in \{\alpha_0, ..., \alpha_{y_i}\}$ for which $\alpha_x = 1$ if $c_{i,t} = x$ and $\alpha_x = 0$ otherwise, capturing the salience of particular contribution amounts, has little effect on the estimated coefficients (see Table C4 in the Appendix). The reported results are also robust when using a probit or a linear probability model instead of a logit model and when using subject fixed effects instead of random effects.

[30] These results continue to hold when we add variables accounting for when subjects contribute according to their own normative contribution vector. In addition, we find that contributions in the category *Overprovided* and *Efficient* are significantly ($p < 0.018$) more stable when subjects are contributing according to their normative contribution vector (see regressions III and IV in Table C4 in the Appendix).

on a contribution vector and thus more difficult to achieve and sustain coordination. Moreover, the stability of contribution vectors, even if they failed to provide the public good, perpetuates coordination failure.

A final open question is whether normative (dis)agreement at the *group level* is correlated with provision success and failure. We note first that, for homogeneous groups, there is perfect normative agreement in 11 of the 15 groups; that is, in these groups all four group members have the same normative contribution vector, which is the unique fairness ideal. In each of the remaining 4 groups, there is only one group member whose normative contribution vector differs from the other three. The importance of normative agreement for successful provision can be seen when comparing these two cases. We find that successful provision in homogeneous groups occurs 86.8% of the time under perfect agreement but only 53.7% of the time when there is disagreement. This is quite a sharp drop considering that there is only one group member who does not agree with the rest in the latter case (see Table C3; Wald test, $p = 0.028$).

In the two heterogeneous treatments, none of the 32 groups exhibit perfect normative agreement. In fact, in 7 groups, there is complete disagreement in that all four group members chose a different normative contribution vector. In 13 groups, there are three different normative contribution vectors, and in 12 groups, there are two. Interestingly, the number of different normative contribution vectors in a group does not impact successful public good provision. Groups with four different normative contribution vectors provide the public good 65.7% of the time, groups with three normative contribution vectors provide 53.1% of the time, and groups with two distinct normative contribution vectors provide 49.2% of the time. These differences are not statistically significant.[31]

It is noteworthy that the relatively low frequency of successful provision in homogeneous groups with normative disagreement(53.7%)is in the realm of the frequency in heterogeneous groups (i.e., 49.2% to 65.7%). This suggests that the relationship between group-level normative disagreement and public good provision is not monotonic. It is plausible that already a very small amount of normative disagreement, that is, when just one group member has a different normative contribution vector, suffices to cause miscoordination and failure to provide the public good. Unfortunately, we cannot test this conjecture because we miss the counterfactual of normative agreement in heterogeneous groups. Another possible explanation for why there is no significant relationship between the amount of normative disagreement within heterogeneous groups and public good provision is that subjects do not know the normative contribution vectors

[31] The statistical tests are based on a logit regression of successful provision on the number of different normative contribution vectors (see Table C3 in the Appendix). Provision rates do not differ depending on the number of normative contribution vectors (Wald tests; $p > 0.132$). This result is robust to considering normative disagreement as measured by the Euclidean distance of the normative contribution vectors (Wald test, $p = 0.235$) and using mean efficiency gains as the outcome variable. Similarly, in homogeneous groups, it is the existence of normative disagreement, and not its magnitude, that explains the drop in provision success (see Table C3).

of their group members. It is difficult, if not impossible, to infer normative judgments based on contribution behavior alone, and subjects may therefore not be able to react to the underlying plurality of normative judgments in a coordinated manner.

5. Conclusion

We have experimentally studied the importance of fairness ideals and normative contribution vectors for the success and failure of efficient coordination. Specifically, we conducted a series of threshold public goods game experiments and compared the cases of homogeneous and two versions of heterogeneous groups with equivalent equilibrium sets. Importantly, with an incentivized task, we elicited subjects' judgments about the normatively desired contribution vectors as well as their expectations about the judgments of others.

Despite there being several thousand efficient Nash equilibrium contribution vectors, in homogeneous groups, subjects largely agree on a unique normative contribution vector (equal contributions), and in heterogeneous groups on three equilibrium contributions (equal contributions, proportional contributions, and contributions leading to equal earnings). In the public goods game itself, we see that, compared to homogeneous groups, heterogeneous groups are much less able to sustain efficient coordination, which is a consequence of the multiplicity of fairness ideals in these groups. Contributions at fairness ideals are the most stable contribution vectors, hence promoting coordination success in homogeneous groups. In contrast, heterogeneous groups frequently fail to coordinate on one of the multiple fairness ideals, leading to instability and coordination failure.

Our results show that fairness ideals can be strong focal points, but also that they facilitate efficient coordination only when they are unique, which is unlikely to be the case in heterogeneous groups. The stability of coordination at fairness ideals and their multiplicity in heterogeneous groups indicate the importance of agreeing on which of them to coordinate. That fairness ideals are also important outside the laboratory and that real actors are able to agree on them can be highlighted by various examples, some of which we described in the introduction.

Until recently, NATO's funding guidelines stipulated that countries should spend 2% of their gross domestic product on their military capabilities, which was increased to 5% at the 2025 NATO Summit in The Hague. This contribution rule is in line with the fairness ideal of proportionality with respect to endowments, which we largely observed in our treatment where players differ in their endowments. Another example of a funding scheme that resembles a fairness ideal is the one used by the World Trade Organization, requiring its members to contribute in relation to their share of international trade, which can be seen as a proxy for the benefits derived from free trade. This fairness ideal is akin to proportional contributions relative to the benefit received in our treatment with heterogeneous benefits from the public good. These are two examples in which the actors could agree on a fairness ideal, ensuring relative stability of the coordination outcome.

However, as we have seen in the experiment, success is not guaranteed, and this challenge is illustrated by the relatively unsuccessful attempts to agree on policies to mitigate climate change. In this realm, global debates about burden sharing, which appeal to different notions of fairness that stem from heterogeneity between countries, have been central in the negotiations at the Conference of the Parties 27 and 28 of the United Nations Framework Convention on Climate Change (Colman and Mathiesen, 2022; Sengupta, 2021; Friedman, 2023). This pattern aligns with our findings on normative disagreement. Yet, we can only speculate about its deleterious effect on the coordination of climate change mitigation efforts, hence the importance of a controlled laboratory experiment.

The plurality of fairness ideals that arise in heterogeneous groups, even in a stylized experimental setting within subjects assigned the same type (i.e., endowment or benefits from the public good), offers an interesting and important avenue for future exploration. How can people reconcile their normative differences to promote socially beneficial outcomes? What mechanisms, institutions, or characteristics of the strategic environment are conducive to successfully resolving normative conflict? Our experiment was not designed to explicitly investigate communication and the back-and-forth process in which subjects can engage in the attainment of efficient coordination. This question may be answered more appropriately within a bargaining framework, where people can engage in making offers and counteroffers or communicate their views to each other to reach a consensus, as it may naturally occur in the field.

References

Adams, J. S. (1965). Inequity in social exchange. In Berkowitz, L., editor, *Advances in Experimental Social Psychology*, volume 2, pages 267–299. Elsevier, Academic Press, Amsterdam.

Alley, R. B., Marotzke, J., Nordhaus, W. D., Overpeck, J. T., Peteet, D. M., Pielke Jr, R. A., Pierrehumbert, R. T., Rhines, P. B., Stocker, T. F., Talley, L. D., et al. (2003). Abrupt climate change. *Science*, 299(5615):2005–2010.

Anderson, L. R., Mellor, J. M., and Milyo, J. (2008). Inequality and public good provision: An experimental analysis. *The Journal of Socio-Economics*, 37:1010–1028.

Avoyan, A. and Ramos, J. (2023). A road to efficiency through communication and commitment. *American Economic Review*, 113(9):2355–2381.

Babcock, L., Issacharoff, S., and Camerer, C. F. (1995). Biased judgments of fairness in bargaining. *American Economic Review*, 85(5):1337–1343.

Bacharach, M. (1993). Variable universe games. In Binmore, K., Kirman, A., and Tani, P., editors, *Frontiers of Game Theory*, pages 255–275. MIT Press, Cambridge, MA.

Bacharach, M. and Bernasconi, M. (1997). The variable frame theory of focal points: An experimental study. *Games and Economic Behavior*, 19:1–45.

Bacharach, M. and Stahl, D. O. (2000). Variable-frame level-n theory. *Games and Economic Behavior*, 32:220–246.

Bagnoli, M. and McKee, M. (1991). Voluntary contribution games: Private provision of public goods. *Economic Inquiry*, 29:351–366.

Bardsley, N., Mehta, J., Starmer, C., and Sugden, R. (2010). Explaining focal points: Cognitive hierarchy theory versus team reasoning. *The Economic Journal*, 120:40–79.

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

Blume, A. and Gneezy, U. (2000). An experimental investigation of optimal learning in coordination games. *Journal of Economic Theory*, 90:161–171.

Bolton, G. E. and Ockenfels, A. (2000). A theory of equity, reciprocity, and competition. *American Economic Review*, 90:166–193.

Brandts, J. and Cooper, D. J. (2006). A change would do you good .... An experimental study on how to overcome coordination failure in organizations. *American Economic Review*, 96(3):669–693.

Buckley, E. and Croson, R. (2006). Income and wealth heterogeneity in the voluntary provision of linear public goods. *Journal of Public Economics*, 90:935–955.

Cappelen, A. W., Hole, A. D., Sørensen, E. Ø., and Tungodden, B. (2007). The pluralism of fairness ideals: An experimental approach. *American Economic Review*, 97(3):818–827.

Cappelen, A. W., Moene, K. O., Sørensen, E. Ø., and Tungodden, B. (2013). Needs versus entitlements—An international fairness experiment. *Journal of the European Economic Association*, 11(3):574–598.

Cappelen, A. W., Mollerstrom, J., Reme, B.-A., and Tungodden, B. (2022). A meritocratic origin of egalitarian behaviour. *The Economic Journal*, 132(646):2101–2117.

Cappelen, A. W., Sørensen, E. Ø., and Tungodden, B. (2010). Responsibility for what? Fairness and individual responsibility. *European Economic Review*, 54(3):429–441.

Cettolin, E. and Riedl, A. (2017). Justice under uncertainty. *Management Science*, 63(11):3739–3759.

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

Colman, Z. and Mathiesen, K. (2022). Cop27 summit yields ‘historic win’ for climate reparationsbut falls short on emissions reductions. *E&E News*.

Cook, K. S. and Hegtvedt, K. A. (1983). Distributive justice, equity, and equality. *Annual Review of Sociology*, 9(1):217–241.

Crawford, V. P., Gneezy, U., and Rottenstreich, Y. (2008). The power of focal points is limited: Even minute payoff asymmetry may yield large coordination failures. *American Economic Review*, 98:1443–1458.

Crawford, V. P. and Haller, H. (1990). Learning how to cooperate: Optimal play in repeated coordination games. *Econometrica*, 58:571–595.

Croson, R. T. A. and Marks, M. B. (2000). Step return in threshold public goods: A meta- and experimental analysis. *Experimental Economics*, 2:239–259.

Erkut, H. and Reuben, E. (2019). Preference measurement and manipulation in experimental economics. In Schram, A. and Ule, A., editors, *Handbook of Research Methods and Applications in Experimental Economics*, chapter 3, pages 39–56. Edward Elgar Publishing, Glos, UK.

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

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

Friedman, L. (2023). Climate summit approves a new fund to help poor countries. *The New York Times*.

Gächter, S. and Riedl, A. (2005). Moral property rights in bargaining with infeasible claims. *Management Science*, 51(2):249–263.

Hamman, J., Rick, S., and Weber, R. A. (2007). Solving coordination failure with "all-or-none" group-level incentives. *Experimental Economics*, 10:285–303.

Hardin, R. (1976). Group provision of step goods. *Behavioral Science*, 21:101–106.

Hirshleifer, J. (1983). From weakest-link to best-shot: The voluntary provision of public goods. *Public Choice*, 41(3):371–386.

Isaac, R. M., Schmitz, D., and Walker, J. M. (1989). The assurance problem in a laboratory market. *Public Choice*, 62:217–236.

Isoni, A., Poulsen, A., Sugden, R., and Tsutsui, K. (2014). Efficiency, equality and labelling: An experimental investigation of focal points in explicit bargaining. *American Economic Review*, 104(10):3256–3287.

Janssen, M. (2000). Rationalizing focal points. *Theory and Decision*, 50:119–148.

Karagözoğlu, E. and Riedl, A. (2015). Performance information, production uncertainty and subjective entitlements in bargaining. *Management Science*, 61(11):2611–2626.

Kendrick, M. S. (1939). The ability-to-pay theory of taxation. *The American Economic Review*, pages 92–101.

Konow, J. (1996). A positive theory of economic fairness. *Journal of Economic Behavior & Organization*, 31(1):13–35.

Konow, J. (2000). Fair shares: Accountability and cognitive dissonance in allocation decisions. *American Economic Review*, 90:1072–1091.

Konow, J. (2003). Which is the fairest one of all? A positive analysis of justice theories. *Journal of Economic Literature*, 41:1186–1237.

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*, 11:495–524.

Ledyard, J. (1995). Public goods: A survey of experimental research. In Kagel, J. H. and Roth, A. E., editors, *Handbook of Experimental Economics*, pages 111–194. rinceton University Press, Princeton, NJ.

Mehta, J., Starmer, C., and Sugden, R. (1994). The nature of salience: An experimental investigation. *American Economic Review*, 84:658–673.

Mellers, B. A. and Baron, J. E. (1993). *Psychological Perspectives on Justice: Theory and Applications.*

Cambridge Series on Judgment and Decision Making. Cambridge University Press.

Mill, J. S. (1871). *Principles of political economy: with some of their applications to social philosophy*, volume 1. Longmans, Green, Reader, and Dyer.

Musgrave, R. A. and Musgrave, P. C. (1989). *Public finance in theory and practice (5th ed.)*. McGraw-Hill, New York.

Nunnari, S. and Pozzi, M. (2022). Meta-analysis of inequality aversion estimates. Working paper no. 9851, CESifo.

Palfrey, T., Rosenthal, H., and Roy, N. (2017). How cheap talk enhances efficiency in threshold public goods games. *Games and Economic Behavior*, 101:234–259.

Peski, M. (2010). Generalized risk dominance and asymmetric dynamics. *Journal of Economic Theory*, 145:216–248.

Rapoport, A. (1988). Provision of step-level public goods: Effects of inequality in resources. *Journal of Personality and Social Psychology*, 54:432–440.

Rapoport, A. and Suleiman, R. (1993). Incremental contribution in step-level public goods games with asymmetric players. *Organizational Behavior and Human Decision Processes*, 55:171–194.

Reuben, E. and Riedl, A. (2013). Enforcement of contribution norms in public good games with heterogeneous populations. *Games and Economic Behavior*, 77:122–137.

Riedl, A., Rohde, I. M., and Strobel, M. (2016). Efficient coordination in weakest-link games. *The Review of Economic Studies*, 83(2):737–767.

Rojo-Arjona, D., Sitzia, S., and Zheng, J. (2022). Overcoming coordination failure in games with focal points: An experimental investigation. *Games and Economic Behavior*, 136:505–523.

Schelling, T. C. (1960). *The Strategy of Conflict*. Harvard University Press, Cambridge, MA.

Schoenbaum, T. J. (2012). *The age of austerity: The global financial crisis and the return to economic growth*. Edward Elgar Publishing.

Sengupta, S. (2021). Calls for climate reparations reach boiling point in glasgow talks. *The New York Times*.

Smith, A. (1937). Innovation and intellectual property rights. In Cannan, E., editor, *An Inquiry into the Nature and Causes of the Wealth of Nations*. Random House, Inc., New York.

Sugden, R. (1993). Thinking as a team: Towards an explanation of nonselfish behaviour. *Social Philosophy and Policy*, 10:69–89.

Sugden, R. (1995). A theory of focal points. *The Economic Journal*, 105:533–550.

Van de Kragt, A. J. C., Orbell, J. M., and Dawes, R. M. (1983). The minimal contributing set as a solution to public goods problems. *The American Political Science Review*, 77:112–122.

Van Huyck, J. B., Battalio, R. C., and Beil, R. O. (1990). Tacit coordination games, strategic uncertainty, and coordination failure. *The American Economic Review*, 80(1):234–248.

Weber, R. A. (2006). Managing growth to achieve efficient coordination in large groups. *American Economic Review*, 96(1):114–126.

Zelmer, J. (2003). Linear public goods experiments: A meta-analysis. *Experimental Economics*, 6:299–310.

Online appendices

Appendix A. Experimental instructions

Below are the instructions for the high type of the *Heterogeneous Benefits* (HB) treatment. The instructions for the *Homogeneous* (H) and *Heterogeneous Endowments* (HE) treatments, as well as those of the low type, are very similar and available upon request.

Welcome

You are now taking part in an economic experiment. Depending on your decisions and the decisions of other participants, you can earn money. How you can earn money is described in these instructions. Therefore, it is important that you read them carefully.

During the experiment, you are not allowed to communicate with other participants in any way. If you have any questions, please raise your hand. We will come to your table to answer your question.

During the experiment, your earnings will be calculated in **tokens**. At the end of the experiment, the total number of tokens you have earned will be converted to money at the following rate: **100 tokens = $1.00**

The following pages describe in detail the experiment.

The Experiment

The experiment is divided into **two independent parts**. In both parts of the experiment, you will make decisions concerning contributions to a group project. Contributions are made in groups of **four** participants. Therefore, all participants in this room have been **randomly** assigned to groups of four. **You have been assigned to be part of group A**. The composition of groups will remain the same throughout the experiment.

We proceed as follows: first, you will be provided with a detailed description of the contribution decision; second, you will be given the instructions and complete part 1 of the experiment; and third, you will be given the instructions and complete part 2 of the experiment.

The contribution decision

Before making the contribution decision, each participant in the group receives **30 tokens**. We will refer to these tokens as the **endowment**. Endowments will be the same throughout both parts of the experiment.

Each group member has to decide how many tokens from their endowment to contribute to the group project. Group members get to keep the tokens they do not contribute. All decisions are made simultaneously. That is, nobody will be informed about the decision of the other group members before everyone makes their decision. The earnings of each group member are the sum of two parts:

1. The number of tokens that each group member kept for themselves.
2. The **income from the group project**.

The income from the group project depends on the **sum of contributions** to the group project by **all** group members. It is determined as follows. If the sum of contributions is **less than 60 tokens**, then the project fails and generates **0 tokens**. If the sum of contributions is **more than or equal to 60 tokens**, then the project succeeds and generates **60 tokens for two group members** and **30 tokens for the other two group members**. In other words, each one of you was **randomly** assigned to receive either 60 or 30 tokens if the group project succeeds. You will be one of the participants who receive **60 tokens if the project succeeds**. The amount each group member receives if the project succeeds will be the same throughout both parts of the experiment.

In summary, after everyone makes their contribution decision, your earnings are equal to **30 – your contribution + (60 if the sum of contributions is more than or equal to 60 tokens)**.

After everyone has made their decision, the contribution stage ends.

Example

Here is an example that illustrates how the earnings are calculated. The numbers used in the example are arbitrary.

A group has four participants (group members 1, 2, 3, and 4). If the group project succeeds, group members 1 and 2 receive 60 tokens, and group members 3 and 4 receive 30 tokens. Group member 1 contributes 18 tokens to the group project, group member 2 contributes 8 tokens, group member 3 contributes 24 tokens, and group member 4 contributes 10 tokens. Since 60 tokens were contributed to the group project in total, earnings are given by:

- Group member 1: $30 - 18 + 60 = 72$ tokens.
- Group member 2: $30 - 8 + 60 = 82$ tokens.
- Group member 3: $30 - 24 + 30 = 36$ tokens.
- Group member 4: $30 - 10 + 30 = 50$ tokens.

Understanding questions

Please answer the following questions.

Question 1: Suppose you contribute 0 tokens to the group project, group member 2 contributes 0 tokens, group member 3 contributes 0 tokens, and group member 4 contributes 0 tokens.

- What are your earnings (in tokens) in this period? ______
- What are member 2's earnings (in tokens) in this period? ______

Question 2: Suppose you contribute 17 tokens to the group project, group member 2 contributes 12 tokens, group member 3 contributes 18 tokens, and group member 4 contributes 14 tokens. [note: contributions randomly generated]

- What are your earnings (in tokens) in this period? ______
- What are member 2's earnings (in tokens) in this period? ______

Question 3: Suppose group member 2 contributes 19 tokens, group member 3 contributes 16 tokens, and group member 4 contributes 11 tokens. [note: contributions randomly generated]

- What are your earnings if you contribute 5 tokens to the project? ______
- What are your earnings if you contribute 15 tokens to the project? ______

Part 1

There are 6 groups of four participants in the room. Recall that you are in group A. In part 1 of the experiment, you will make decisions concerning group B.

Specifically, your task in part 1 is to put yourself in the position of a **neutral uninvolved arbitrator** and indicate the **appropriate contribution** each member of group B **should** make to their group project.

Your decision will determine the earnings of group B in part 1, so please consider your choices carefully. Note that you will not interact with members of group B at any point in the experiment. Moreover, the choices of group B will not affect your earnings in either part 1 or part 2 of the experiment.

Decision in part 1

Please put yourself in the position of a **neutral uninvolved arbitrator** and indicate above the **appropriate contribution** each member of group B **should** make.

- Group member B1: ______
- Group member B2: ______
- Group member B3: ______
- Group member B4: ______

To observe how earnings are affected by contribution decisions, click on **calculate**. Once you click on **submit**, you won't be able to change your answers further.

Part 1: Additional questions

Before we proceed with part 2, we would like you to answer the following questions. Just like you, each member of **your group (i.e., group A)** was asked to indicate the appropriate contribution that each group member in group B should make to the group project.

Next, we want you to **guess the answers submitted by your fellow group members**. You will be remunerated for the accuracy of your guesses. Specifically, the computer will randomly pick one of your guesses and compare it to the corresponding group member's actual decision. If your guess is correct, you earn 200 additional tokens; otherwise, you earn 0 additional tokens. You will be informed of the accuracy of your guess at the end of the experiment.

Guesses for part 1

Group member A2 has an endowment of 30 tokens and receives **60 tokens** if the project succeeds. Please specify below the contributions you expect A2 chose for group B.

- Group member B1: ______
- Group member B2: ______
- Group member B3: ______
- Group member B4: ______

Group member A3 has an endowment of 30 tokens and receives **30 tokens** if the project succeeds. Please specify below the contributions you expect A3 chose for group B.

- Group member B1: ______
- Group member B2: ______
- Group member B3: ______
- Group member B4: ______

Group member A4 has an endowment of 30 tokens and receives **30 tokens** if the project succeeds. Please specify below the contributions you expect A4 chose for group B.

- Group member B1: ______
- Group member B2: ______
- Group member B3: ______
- Group member B4: ______

Part 2

In part 2 of the experiment, you will interact for **20 periods** with participants in group A. In every period, you and your other group members make a contribution decision, after which you will be informed of the outcome of that period. Remember that group composition remains the same for the entire experiment. Similarly, your endowment of 30 tokens and the 60 tokens you receive if the project succeeds will remain the same for all 20 periods.

Click **ready** to start part 2 of the experiment.

Decisions in part 2

How many tokens do you want to contribute to the group project? ______

Appendix B. Generalized risk dominance

In this section, we draw on a generalized notion of risk dominance developed by Peski (2010) to demonstrate the absence of an obvious risk ordering among the equilibria of our game that are on the Pareto frontier. Peski (2010) develops two notions of *generalized risk dominance* ("GRD"), one ordinal and one cardinal. The ordinal definition does not require existence or specification of a utility function; cardinal GRD, hence, not surprisingly, implies ordinal GRD. We shall proceed by showing that none of the equilibria on the Pareto frontier satisfies ordinal GRD (and hence none satisfies cardinal GRD). But first, some definitions.

For details on the notation and definitions, the reader should be familiar with the notation of Peski (2010), as we do not repeat it here for succinctness.

Fix an allocation $\sigma = (c_1, c_2, c_3, c_4) \in \Sigma$. The ordered pair $(\eta, \bar{\eta}) \in \Sigma^2$ is *σ-associated* if, for all players i, either $\eta_i = c_i$ or $\bar{\eta}_i = c_i$. Allocation σ is *ordinal generalized risk-dominant (ordinal GRD)* if, for all σ-associated profiles η and $\bar{\eta}$, for all players i, action c_i is a best response to either η or η_i. In other words, σ is an ordinal GRD if, whenever c_i is not a best response against η, it is a best response against *any* σ-associated profile $\bar{\eta}$.

Let $u_i(x_i, \eta)$ be the payoff of player i when she plays x_i and the other players follow profile η. Then profile σ is *cardinal generalized risk-dominant (cardinal GRD)* if, for all σ-associated profiles η and $\bar{\eta}$, for all players i, $\max_{x_i \neq c_i} u_i(x_i, \eta) - u_i(c_i, \eta) \leq u_i(c_i, \bar{\eta}) - \max_{x_i \neq c_i} u_i(x_i, \bar{\eta})$

Claim. *There is no $\sigma \in \Sigma \backslash z$ that satisfies ordinal GRD, for any of the parameter configurations that we consider in our treatments.*

Proof: Consider any $\sigma \in \Sigma \backslash z$. Then we know $\sum_j c_j = 60$, and $0 \leq c_j \leq 30$ for all j. We distinguish two cases:

(i) $\min\{c_1, c_2, c_3, c_4\} > 0$: Consider $\eta = (c_1, c_2, 0, 0)$ and $\bar{\eta} = (0, 0, c_3, c_4)$. Suppose $c_1 + c_2 < 30$. Then neither c_1 nor c_2 can be best replies to η. Yet they cannot both be best replies to $\bar{\eta}$, since that would imply $c_1 + c_3 + c_4 = 60$ and $c_2 + c_3 + c_4 = 60$, which given $\min\{c_1, c_2, c_3, c_4\} > 0$ contradicts $\sum_j c_j = 60$. The case $c_1 + c_2 > 30$ is analogous. Suppose finally $c_1 + c_2 = c_3 + c_4 = 30$. Then $\min\{c_1, c_2, c_3, c_4\} > 0$ implies $\max\{c_1, c_2, c_3, c_4\} < 30$. But then all best replies to η and $\bar{\eta}$ are either 30 or 0. Apply the definition to conclude that σ is not ordinal GRD.

(ii) $\min\{c_1, c_2, c_3, c_4\} = 0$ We can have $c_i = 0$ for at most two players. Assume first w.l.o.g. that $c_4 = 0$ and $\min\{c_1, c_2, c_3\} > 0$. Let $\eta = (c_1, c_2, 0, 0)$ and $\bar{\eta} = (0, 0, c_3, 0)$. Clearly neither c_1 nor c_2 can be best replies to η since $c_1 + c_2 < 60$ by assumption. But being strictly positive they cannot be best replies to $\bar{\eta}$ either since $c_3 \leq 30$ and $\max\{c_1, c_2\} < 30$. Assume now w.l.o.g. $c_3 = c_4 = 0$. Then $c_1 = c_2 = 30$. Let $\eta = (c_1, 0, c_3, c_4)$ and $\bar{\eta} = (0, c_2, 15, 0)$. Then c_1 is neither a best reply to η nor to $\bar{\eta}$. This concludes the proof of the claim. ■

Corollary. *There is no $\sigma \in \Sigma \backslash z$ that satisfies cardinal GRD, for any of the parameter configurations that we consider in our treatments.*

Appendix C. Additional tables

Table C1. Additional descriptive statistics of the subjects' own and expected normative contribution vectors in heterogeneous groups by type

Note: Fraction of normative contribution vectors that are: Successful ($C \geq \Gamma$), Efficient ($C = \Gamma$ and $b_i\Gamma \leq c_i \, \forall i$), and a Fairness ideal.

	i's normative judgment				**i's expected normative judgment of j**							
i's type	Low		High		Low				High			
j's type					Low		High		Low		High	
	HE	**HB**	**HE**	**HB**	**HE**	**HB**	**HE**	**HB**	**HE**	**HB**	**HE**	**HB**
Successful	1.000	1.000	0.969	1.000	1.000	1.000	0.969	0.953	0.953	0.953	1.000	0.969
Efficient	0.969	0.938	0.938	1.000	0.969	0.938	0.953	0.938	0.938	0.953	0.938	0.969
Fairness ideal	0.750	0.719	0.688	0.813	0.781	0.844	0.797	0.766	0.688	0.813	0.781	0.906

Table C2. Contribution vector in period $t+1$ depending on the contribution vector in period t

Note: Multinomial logit model with the contribution vector played in period $t+1$ as the dependent variable and dummy variables indicating the contribution vector played in period t as independent variables. Contribution vectors are classified as either *Failed* ($C_t < \Gamma$), *Overprovided* ($C_t > \Gamma$), *Efficient* ($C_t = \Gamma$ and $b_i\Gamma \leq c_{i,t}\,\forall i$, excluding fairness ideals), or *Fairness ideal.* Independent variables are interacted with a dummy variable indicating whether a group is homogeneous or heterogeneous. The table reports marginal effects calculated separately for each treatment with *Failed* contribution vectors as the omitted category (the constant). Robust standard errors clustered at the group level are in parentheses. ** and * denote statistical significance at 1% and 5%, respectively. The regression is based on 893 observations from 47 groups. The resulting log pseudolikelihood equals -666.821 and the pseudo-R^2 is 0.414.

Panel A. Homogeneous groups

		Period $t+1$			
		Failed	Overprovided	Efficient	Fairness ideal
Period t	Overprovided	-0.527^{**}	0.129	0.058	0.340^{**}
		(0.140)	(0.103)	(0.058)	(0.124)
	Efficient	-0.454^{**}	0.102	0.467^{**}	-0.115^{**}
		(0.111)	(0.125)	(0.065)	(0.060)
	Fairness ideal	-0.712^{**}	-0.072^{*}	-0.033	0.818^{**}
		(0.076)	(0.036)	(0.018)	(0.069)
	Constant	0.754^{**}	0.098^{**}	0.033^{**}	0.115^{**}
		(0.075)	(0.034)	(0.018)	(0.060)

Panel B. Heterogeneous groups

		Period $t+1$			
		Failed	Overprovided	Efficient	Fairness ideal
Period t	Overprovided	-0.324^{**}	0.245^{**}	0.063	0.016
		(0.069)	(0.061)	(0.037)	(0.018)
	Efficient	-0.444^{**}	-0.038	0.514^{**}	-0.032^{**}
		(0.089)	(0.059)	(0.102)	(0.010)
	Fairness ideal	-0.621^{**}	-0.185^{**}	-0.068^{**}	0.875^{**}
		(0.050)	(0.042)	(0.019)	(0.037)
	Constant	0.687^{**}	0.214^{**}	0.068^{**}	0.032^{**}
		(0.046)	(0.037)	(0.019)	(0.001)

Table C3. Successful provision depending on the degree of normative disagreement

Note: Logit regression of successful provision of the public good as the dependent variable in columns I, II, V, and VI. Linear regressions of group efficiency gains as the dependent variable in columns II, IV, VI, and VIII. As independent variables, columns I, III, V, and VII include dummy variables indicating the number of different normative contribution vectors in a group. In columns I and II, the base level is 2 different normative contribution vectors, while in columns V and VI, it is 1. Columns II, IV, VI, and VIII include the Euclidean distance between the normative contribution vectors in a group as the dependent variable. The table reports marginal effects and robust standard errors clustered at the group level in parentheses. ** and * denote statistical significance at 1% and 5%, respectively.

	Heterogeneous groups				Homogeneous groups			
	Provision		Efficiency gain		Provision		Efficiency gain	
	I	II	III	IV	V	VI	VII	VIII
2 different normative contribution vectors					−0.331*		−0.580*	
					(0.151)		(0.237)	
3 different normative contribution vectors	0.039		0.068					
	(0.118)		(0.151)					
4 different normative contribution vectors	0.165		0.101					
	(0.107)		(0.158)					
Normative disagreement (Euclidean distance)		0.017		0.019		−0.011		−0.027
		(0.017)		(0.022)		(0.019)		(0.045)
Constant	0.492**	0.411*	0.259**	0.163	0.868**	0.799**	0.799**	0.682**
	(0.077)	(0.161)	(0.093)	(0.201)	(0.074)	(0.077)	(0.094)	(0.108)
Number of observations	640	640	640	640	300	300	300	300
R^2 / Pseudo-R^2	0.012	0.006	0.003	0.009	0.108	0.008	0.145	0.014

Table C4. Stability of individual contributions (probability that $c_{i,t} = c_{i,t+1}$)

Note: Logit regressions with subject random effects. The dependent variable equals one if $c_{i,t} = c_{i,t+1}$ and zero otherwise. Contribution vectors are classified as either *Failed* ($C_t < \Gamma$), *Overprovided* ($C_t > \Gamma$), *Efficient* ($C_t = \Gamma$ and $b_i\Gamma \leq c_{i,t}\, \forall i$, excluding fairness ideals), or *Fairness ideal.* All regressions contain 2432 observations from 128 subjects in 32 groups. The table reports marginal effects and robust standard errors clustered at the group level in parentheses. ** and * denote statistical significance at 1% and 5%, respectively.

	I	II	III	IV
Failed × $c_{i,t}$ is a fairness ideal		0.129**	0.098*	0.131**
		(0.041)	(0.043)	(0.044)
Failed × $c_{i,t}$ is i's normative contribution			−0.003	−0.006
			(0.057)	(0.051)
Overprovided	0.073**	0.063	−0.008	0.054
	(0.026)	(0.051)	(0.051)	(0.051)
Overprovided × $c_{i,t}$ is a fairness ideal		0.142*	0.079	0.102
		(0.064)	(0.057)	(0.070)
Overprovided × $c_{i,t}$ is i's normative contribution			0.176**	0.146**
			(0.053)	(0.055)
Efficient	0.241**	0.272**	0.268*	0.259**
	(0.042)	(0.074)	(0.108)	(0.081)
Efficient × $c_{i,t}$ is a fairness ideal		0.074	−0.038	0.026
		(0.077)	(0.102)	(0.076)
Efficient × $c_{i,t}$ is i's normative contribution			0.161*	0.148*
			(0.074)	(0.069)
Fairness ideal	0.440**	0.529**	0.497**	0.521**
	(0.029)	(0.040)	(0.050)	(0.047)
Fairness ideal × $c_{i,t}$ is i's normative contribution			0.038	0.016
			(0.045)	(0.031)
Constant	0.516**	0.436**	0.463**	0.436**
	(0.025)	(0.036)	(0.037)	(0.036)
Controls	Yes	Yes	No	Yes
Pseudo-R^2	0.132	0.136	0.099	0.140