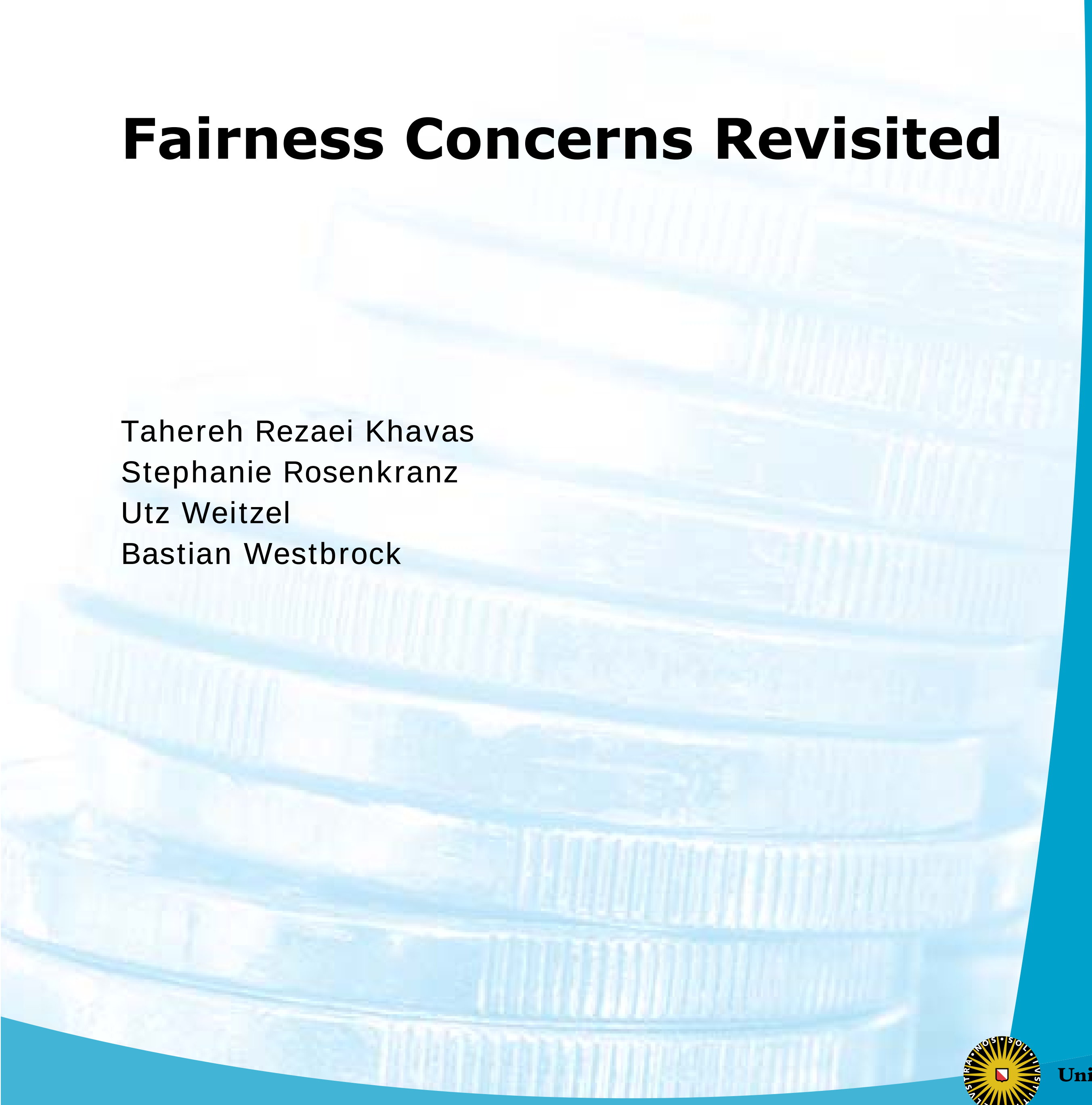
Tjalling C. Koopmans Research Institute

Tjalling C. Koopmans

Discussion Paper Series nr: 14-13



Fairness Concerns Revisited

Tahereh Rezaei Khavas
Stephanie Rosenkranz
Utz Weitzel
Bastian Westbrock

Universiteit Utrecht

Utrecht School of Economics

Tjalling C. Koopmans Research Institute
Utrecht School of Economics
Utrecht University

Kriekenpitplein 21-22
3584 EC Utrecht
The Netherlands
telephone +31 30 253 9800
fax +31 30 253 7373
website www.koopmansinstitute.uu.nl

The Tjalling C. Koopmans Institute is the research institute and research school of Utrecht School of Economics.
It was founded in 2003, and named after Professor Tjalling C. Koopmans, Dutch-born Nobel Prize laureate in economics of 1975.

In the discussion papers series the Koopmans Institute publishes results of ongoing research for early dissemination of research results, and to enhance discussion with colleagues.

Please send any comments and suggestions on the Koopmans institute, or this series to J.M.vanDort@uu.nl

ontwerp voorblad: WRIK Utrecht

How to reach the authors

Please direct all correspondence to the last author.

Tahereh Rezaei Khavas*
Stephanie Rosenkranz*
Utz Weitzel#
Bastian Westbrock*
*Utrecht University
Utrecht School of Economics
Kriekenpitplein 21-22
3584 TC Utrecht
The Netherlands.
E-mail: B.Westbrock@uu.nl
#Radboud Universiteit Nijmegen
Faculteit der Managementwetenschappen
Postbus 9108
6500 HK Nijmegen

This paper can be downloaded at: http:// www.uu.nl/rebo/economie/discussionpapers

Utrecht School of Economics
Tjalling C. Koopmans Research Institute
Discussion Paper Series 14-13

FAIRNESS CONCERNS REVISITED

Tahereh Rezaei Khavas[a]
Stephanie Rosenkranz[a]
Utz Weitzel[b]
Bastian Westbrock[a]

[a] Utrecht School of Economics
Utrecht University

[b] Nijmegen School of Management
Radboud University Nijmegen

September 2014

Abstract

Experimental economics has provided evidence for fairness concerns, but their relative strength and even their stability is still under debate. We reconcile the seemingly inconsistent results by presenting a theory of marginal fairness concerns. The key assumption is that fairness concerns are stable across various decision situations, but individuals care only marginally about other individuals' payoffs. This produces inequitable outcomes when the decision situation is 'unfair' but equitable outcomes when the structure itself is 'fair'. An experimental horse race with competing theories of pure selfishness, pure fairness, and power-/need-based norms, applied across a range of (a)symmetric and (in)transitive experimental decision settings, supports our theory: 80% of the subjects in our experiment appear to be at most marginally fairness concerned.

Keywords: Fairness, Lab experiment, Local Public Goods Game, Heterogeneous Influence Structure.

JEL classification: C91, C70, H41, D85

1 Introduction

While a long-standing tradition in economics views humans as being exclusively self-interested, experimental economics has repeatedly shown that they often deviate from purely self-interested behavior in a fair and reciprocal manner. In laboratory experiments on bargaining and cooperation, a large percentage of individuals have been found to exhibit other-regarding behavior "that the self-interest hypothesis cannot rationalize in any reasonable way."[1] At the same time, there is ample experimental evidence (e.g., Smith, 1962, Davis and Holt, 1993, Plott, 1983) that "other-regarding motives only have a limited impact on behavior and that the self-interest assumption provides a good description for most people's behavior."[2]

These seemingly inconsistent results convey the impression that we have to let go of a parsimonious theory to make sense of individual behavior in the lab. Along these lines, some experimenters interpret their findings as to reflect unstable social preferences that are switched on and off depending on the specific decision situation in the lab (see, e.g., List, 2007 and the discussion in Rabin and Charness, 2002, and Sobel, 2005). Theories of power-based norms, for example, assert that an individual only feels envy if another person in a less powerful position earns more than her but not if the other person is more powerful. Other experimenters assume that social preferences are stable but whether they are displayed or not depends on the details of the decision situation in the lab.[3] Fehr and Schmidt (2006) and Fehr and Fischbacher (2002), for example, argue that market competition can make the achievement of other-regarding goals infinitely costly. There is, however, no theory to explain the relative costs, nor a systematic analysis of how these costs could lead to the diversity of observed outcomes.

The present study proposes and experimentally evaluates a parsimonious utility model to reconcile the stylized facts across a systematic variety of decision contexts. We expand on the class of 'difference aversion models' where – next to an individual's own payoffs – the (relative) payoffs of others enter an individual's utility function. Our key assumption is that all individuals are fairness concerned in all decision situations but only to a small extent or, in economics jargon, only marginally. Individuals pay most attention to their own payoffs. Their behavior therefore largely resembles selfish best response play. They only deviate in favor of a more fair outcome in the vicinity around their best response option where the monetary losses of deviation are small. Thus, the main feature of this theory of marginal fairness concerns is that its predictions closely resemble selfish equilibrium play whereby individuals deviate in a systematic manner from their best responses.

First, a possibly surprising finding of our paper is that marginal fairness concerns produce a fine-grained selection on the set of equilibria when there

[1] See, e.g., the survey on "economics of fairness, reciprocity and altruism" by Fehr and Schmidt (2006, p.615).

[2] Idem.

[3] Andreoni and Miller (2002) explicitly test for the stability of other-regarding preferences.

are multiple elements in this set. In this regard, our model bears a close resemblance with prior theories that view fairness as an equilibrium selection device (Binmore, 2005). Second, marginal fairness concerns are able to explain much of the behavioral variation found in previous experimental studies. The vast majority of experiments in support of strong other-regarding preferences consider decision situations in which the actions of all individuals affect the payoffs of *all* other individuals (i.e., transitive decision situations), and mostly also to the same extent (i.e., symmetric situations). Examples are the early studies by, e.g., Güth, Schmittberger, and Schwarze (1982), Forsythe, Horowitz, Savin, and Sefton (1994), Marwell and Ames (1980), Marwell and Ames (1981), Smith (1979), Smith (1980), Isaac and Walker (1988b), and Roth, Prasnikar, Fujiwara, and Zamir (1991).[4] Evidence against the importance of other-regarding preferences, on the other hand, comes from experiments in which individuals affect other individuals' profits in a heterogeneous way, such as in multilateral bargaining (Roth, Prasnikar, Fujiwara, and Zamir, 1991), sequential best-shot public goods games (e.g. Harrison and Hirshleifer, 1989), or market games (Plott, 1982). These studies suggest that the asymmetry of the influence structure moderates the role of other-regarding preferences for behavior.

In line with these studies, we investigate decision situations where individual decisions have an asymmetric and/or intransitive influence on other individuals' payoffs and compare them with decision making in completely homogeneous (symmetric and transitive) influence structures. In particular, we derive and experimentally assess the predictions of our marginal fairness theory for a (local) public goods game (Bramoullé and Kranton, 2007). The imperfectness of the influence structures in this game implies that decisions impose local (positive) externalities. In asymmetric settings, some individuals have more interaction partners than others and can therefore obtain more public goods. In intransitive settings, for example, where individuals A and B as well as A and C share with each other but individuals B and C do not, only a subgroup of the individuals benefit from someone's public good contribution.

Based on a public good game with interior equilibria, we derive predictions for our marginal fairness model regarding fair equilibrium play. We compare the predictions with those of several prominent, alternative fairness theories, such as difference aversion with strong fairness concerns, a social welfare model (Rabin and Charness, 2002), and two difference aversion models with unstable fairness concerns. For the latter we assume that an individual's preference for fairness depends on the influence structure, i.e., the asymmetry of the structure itself establishes a distributive norm (e.g., power or need), which is then reflected in the individual's preferences. In all these models, we allow for (some) individual heterogeneity in fairness concerns to acknowledge available experimental evidence suggesting that many individuals behave quite selfishly in various circumstances (Fehr and Schmidt, 1999). Finally, we compare the

[4] More recent experiments with homogeneous decision situations are Blanco, Engelmann, and Normann (2011) and Fisman, Kariv, and Markovits (2007) on bilateral bargaining, Fehr and Falk (1999) on double auctions, and Andreoni and Miller (2002), and Andreoni, Harbaugh, and Vesterlund (2003) on cooperation for public goods provision.

predictions with the equilibrium contributions of purely self-interested individuals.

We then run a horse race between these models by experimentally evaluating their predictions for the local public goods games. Our experimental evidence contradicts theories asserting strong other-regarding fairness or altruistic motives for a considerable share of individuals. In addition, it refutes the predictions of the theory of need-based norms. Instead, the best description of our findings is provided by the model of marginal fairness concerns.

Marginal fairness concerns are sufficient to produce outcomes looking remarkably fair when the interaction structure itself is 'fair' but rather selfish outcomes when the structure is 'unfair'. The intuition is that in symmetric and transitive structures, marginally fairness concerned individuals are predicted to choose contribution levels so that the total is close to the payoff-maximum level. However, on this domain of the public goods payoff function, an incremental change in an individual's contribution level is not (very) payoff relevant to her. Hence, she may be willing to make an effort to equalize payoff differences: an individual contributing less than the rest may want to contribute more, while an individual who over-contributes may want to cut back on contributions. In equilibrium, contributions need to be equalized. In asymmetric structures, on the other hand, marginal fairness theory predicts individuals in more influential positions will free ride because what they receive from their neighbors by far exceeds their individually desired contribution level, and the achievement of a fairer outcome comes at a high marginal cost to them. The predicted behavior therefore bears a close resemblance with purely selfish play.

Another theory with similar predictive power for our experimental findings on asymmetric structures is a power-based fairness norm. Marginal fairness theory has, however, several advantages. First, the theory is more parsimonious as it is based on stable social preferences. Second, despite its simplicity, the theory is better able to explain our experimental findings on a specific class of symmetric but intransitive interaction structures. Here, some groups of individuals opted for roughly equitable contribution levels, while other groups supported free riders in their midst. However, as there are ex ante no criteria why some individuals in a symmetric structure should feel entitled to earn more than others, only the former type of play can be easily reconciled with a power-based norm. Both types of equilibria can, however, be explained by marginal fairness theory. Third, a consistent finding of all our experimental settings is that the vast majority of individuals invest approximately in line with their selfish best response contribution level. This only makes sense from the perspective of stable and weak other-regarding concerns. In fact, our experimental design with decreasing marginal per capita returns allows us to interpret deviations from best response play as individual costs for other-regarding behavior. Our findings suggest that roughly 80% of the individuals in our experiments are at most marginally concerned about others' payoffs.

Overall, we can conclude that by considering mainly transitive and symmetric settings in experimental studies the effect of fairness concerns on individual decision making is potentially overrated. These settings make it difficult to distinguish between strong and merely marginal fairness concerns. Payoff equality is chosen by all individuals who are not purely selfish, even if they care only marginally about other individuals' payoffs. This problem applies in particular to linear payoff games such as common bargaining games or threshold public goods environments where even small fairness concerns might lead to observations of completely fair solutions. Another overarching conclusion of our study is that, for asymmetric influence structures, fairness concerns are no remedy against the existence of rather unfair payoff distribution. Quite to the contrary, our theoretical predictions and experimental findings show, in terms of our highly asymmetric and intransitive influence structures, that typically the most unfair outcomes are selected.

Consistent with our theory and evidence on marginal fairness concerns, several empirical studies indicate that pro-social behavior is less frequent in settings characterized by intransitivity and/or asymmetry. The sociological literature on power in exchange networks finds similar negative relations between pro-social behavior and network centrality (see, e.g., Cook and Emerson, 1978, Lovaglia, Skvoretz, Willer, and Markovsky, 1995 or Molm, Peterson, and Takahashi, 1999) and between social cohesion and network centrality (Lawler and Yoon, 1998). Studies by Joung, Chiu, and Chen (2012) and Ramzan, Park, and Izquierdo (2012) suggest free riding to be more common in peer-to-peer (P2P) networks, i.e., social networks for pooling information resources (videos, files, etc.), which are organized as fixed structures (i.e., trees) or as centralized (unstructured) networks, than in decentralized (or purely unstructured) P2P networks. In contrast to the more star-like P2P interaction, the latter structure consists of peers that are all equal, resulting in a fully transitive and symmetric flat overlay network, where fairness concerns, even when only marginal, seem to be sufficient to select more equitable outcomes with less free riding.

The remainder of the paper is organized as follows: In the next section, we briefly characterize the incentives and interaction structures on which we base and develop our theory and which we implement in the experiment. In Section 3 we present our theoretical predictions. In Section 4 we describe the experimental setup. Our results are presented in Section 5. In Section 6 we discuss the robustness of our findings and in Section 7 we conclude.

2 The local public goods game

To specify our theoretical predictions and experimental tests, we study a game introduced by Bramoullé and Kranton (2007), where players jointly invest in the production of a (local) public good.[5] The specific feature of this game is that the size of the public good received by a player does not depend on every player's contribution but only on that of the player's neighbors in a

[5] Following the convention in the game theoretical literature we refer to individuals as players in our analysis of the local public goods game and when we discuss theoretical predictions. When we refer to the participants in our experiment, following the convention in the experimental literature, we use the term 'subjects'.

predefined interaction structure. Specifically, a player's payoff is given by:

$$\pi_i = b_i - ce_i = b\big(e_i + \sum_{j \in N_i(g)} e_j\big) - ce_i \tag{1}$$

where e_i denotes player i's contribution to the public good and $N_i(g)$ denotes the set of player i's neighbors in an interaction structure g. Moreover, Bramoullé and Kranton (2007) assume that the benefit function $b(\cdot)$ is increasing and concave in contribution levels, i.e. $\frac{db}{de_j} > 0$, $\frac{d^2b}{d^2e_j} < 0$ for any $j \in N_i(g) \cup i$. This implies that the *marginal material per capita return* is not constant, such as in typical public goods games with linear production functions, but depends on the total contributions in a player's neighborhood. Also, the payoff-maximizing contribution is positive even for a single player, $b'(0) - c > 0$, which implies that there is some privately optimal contribution level up to which a selfish player would invest.

These properties are desirable for our purposes, because they allow us to evaluate the intensity of an individual's fairness concerns on an interval scale. This is difficult in linear contribution or threshold public goods environments because any existing difference aversion model would predict corner solutions: either the marginal monetary losses are lower than the marginal benefits from a more equitable, moral, or efficient outcome, in which case utility is maximized by choosing the most pro-social contribution, or the monetary losses are larger so that a (zero) contribution is chosen out of monetary concerns. In a nonlinear environment, however, positive (interior) contribution levels are optimal even for selfish individuals, different degrees of other-regarding concerns are expected to lead to stronger or weaker deviations from the selfish optimum.[6]

For illustration, and as a setup for our experimental evaluation, Figure 1 shows all connected interaction structures with four players (plus the dyad). The structures can be ranked in terms of two properties: their degree of transitivity and symmetry. The dyad and the complete interaction structure, i.e., the two structures investigated in many prior 2-player respectively 4-player public goods games, including the circle, are fully symmetric. All players have an equal number of neighbors. The star, on the other hand, is highly asymmetric. The transitivity measures the number of complete triplets in a structure, i.e., the sets of three players such that if i and j as well as j and k are neighbors, then i should also be in the neighborhood of k. The complete interaction structure is fully transitive. The star and the circle, on the other hand, have zero transitivity.

(Figure 1 about here)

An intriguing, feature of the Bramoullé and Kranton (2007) local public goods game is that certain interaction structures potentially produce considerable inequality in the players' access to others' contributions. For instance, in

[6] Symmetric public goods games with interior equilibria have been studied e.g. by Sefton and Steinberg (1996), Laury and Holt (1998), Isaac and Walker (1998), Willinger and Ziegelmeyer (2001), and Laury, Walker, and Williams (1999).

the star the center player has access to every other player, while the peripheral players only interact with the center player. Furthermore, the lack of transitivity in the circle might result in distinct levels of the public good provided in different neighborhoods. But, the ultimate cause for any payoff inequality lies in the players' reluctance to contribute above (or to fall short of) their payoff maximizing contribution level. The game therefore provides a rich test bed for fairness concerns. At the same time, it allows testing for "social welfare" concerns as the contribution level maximizing the group payoff exceeds the individual optimum in every interaction structure.

In the following, we develop and test predictions for several prevalent theories of other-regarding behavior concerning the dyad, complete, star, and circle interaction structures. We then experimentally evaluate these predictions and perform robustness checks on the remaining three interaction structures (line, dbox, core-periphery).

3 Theoretical predictions

In the economics literature the common ground of most fairness theories is that other players' payoffs enter an individual's utility function.[7] We follow this tradition and assume that the subjects in our experiment do not only take the payoff function (1) into consideration but also the payoffs of others in their neighborhood (or even of all other players in the same game).

Expanding on the hypothesis about distributional preferences being expressed by inequity or difference aversion, as exemplified by Loewenstein, Thompson, and Bazerman (1989), Bolton and Ockenfels (2000), and Fehr and Schmidt (1999), we let fairness concerns enter a player's utility in the following form:

$$U_i = \pi_i - \sum_{j \in \overline{N}_i(g)} \alpha_i [\pi_j - \pi_i] - \sum_{j \in \underline{N}_i(g)} \beta_i [\pi_i - \pi_j] \tag{2}$$

Envy and guilt, the two faces of inequity aversion, are introduced in the following form: envy is represented by the first subtrahend. $\overline{N}_i(g)$ denotes the set of player i's neighbors who earn more than her ($\pi_j - \pi_i > 0$) and α_i reflects the player- and context-specific intensity of envy. The sense of guilt is dependent on $\underline{N}_i(g)$, the number of neighbors who earn less ($\pi_j - \pi_i < 0$), and β_i, the guilt parameter.

Utility function (2) allows for a large degree of heterogeneity across players and positions in an interaction structure. Not only may different players experience varying intensities of envy or guilt, but their perception of fairness may

[7] The literature distinguishes two strands of theories: 'outcome-based' and 'intention-based' theories of fairness concerns. The 'outcome-based' theories assume that players are only concerned about the distributional consequences of their actions but not about the intentions driving their opponents to choose these actions. The latter was suggested by Rabin (1993), who started from the observation that behavior is often a reaction to the (expected) intentions of other people. The theory predicts that individuals do not undertake kind actions unless others have shown kind intentions. Given our one-shot interaction structure and the difficulties in measuring beliefs, we confined our analysis to outcome-based models.

also be position-dependent in the sense that the asymmetry of the structure itself establishes a distributive norm (e.g., power or need). This renders a comprehensive set of predictions difficult. In the first instance, we therefore confine our predictions to players who differ in their fairness concerns only to a limited extent. In other words, put in the same position in an interaction structure and confronted with the same neighborhood contributions, all players would respond in a similar way. This is backed by theories regarding fairness concerns as a norm shared by all members of the same society (Binmore, 2005). We defer our discussion of substantial player heterogeneity to Section 6.1.[8]

3.1 Pure selfishness

Bramoullé and Kranton (2007) were the first to study equilibrium play in the local public goods game. Their predictions concern a group of entirely selfish players. They show that in order to maximize payoff function (1) an egoistic player responds to the neighbors' contributions by choosing e_i such that:

1. $e_i = e^* - \sum_{j \in N_i(g)} e_j$ if and only if $\sum_{j \in N_i(g)} e_j \leq e^*$
2. $e_i = 0$ if and only if $\sum_{j \in N_i(g)} e_j \geq e^*$

where e^* is uniquely defined by $b'(e^*) - c = 0$. Hence, there exists a maximum contribution level e^* in every neighborhood of the game, and players aim to fill the gap between this value and their neighbors' contributions. If, however, the latter already exceeds e^*, then players free ride. We use e^* as a benchmark for comparison throughout.

The best response behavior specified above results in equilibrium profiles characterized in Bramoullé and Kranton (2007), which depend on the precise interaction structure. The equilibria for the four interaction structures – dyad, complete, star, and circle – are summarized in Table 1.

3.2 Pure fair play

At the other extreme, all players might be driven by pure fairness concerns (with some limited degree of variability). In terms of utility function (2), this can be represented by large parameter values for α_i and β_i, rendering player's concerns for own payoff consequences negligible. A fair contribution profile must therefore equalize payoffs for all players i and j in the game:

$$b_i - ce_i = b_j - ce_j$$

[8] We also make two other key assumptions as follows: according to utility function (2), a player compares herself to every neighbor. Thus, her envy or guilt increases with neighborhood size. Note, however, that our following predictions are perfectly compatible with parameters α_i and β_i, which shrink (or even increase) with neighborhood size. Second, a player compares herself only to her direct neighbors and disregards 'social welfare' concerns. In Section 6.3, we argue that the predictions are quite similar when players compare themselves to everyone else in the game and that such a theory is equally well backed by our experimental findings. Moreover, we will see that 'social welfare' does not play a major role in our experiment.

For some of the interaction structures in our experiment, we can additionally exploit the following result:

Proposition 1. *Suppose players are driven by pure fairness concerns. If $N_j(g) \subset N_i(g)$, then $e_j \leq e_i$, whereas if $N_j(g) = N_i(g)$, then $e_j = e_i$.*

Proof. Suppose two players with $N_j(g) \subset N_i(g)$. Then $b_i = b(e_1 + e_2 + ... + e_p + ... + e_q)$ and $b_j = b(e_1 + e_2 + ... + e_p)$. Clearly, $b_i > b_j$ if for at least one player q, where $q \in N_i(g) \backslash N_j(g)$, $e_q > 0$. Hence, $e_i > e_j$ should hold in order to have $\pi_i = \pi_j$. On the other hand, if $e_q = 0$ for all players q, then $b_i = b_j$ and also $e_i = e_j$. Finally, suppose $N_j(g) = N_i(g)$. Then $b_i = b_j$ and $e_i = e_j$. □

The prediction is relevant for the complete interaction structure and the dyad, where every profile is fair as long as all players contribute the same amount. This also includes some extreme profiles, e.g., where nobody contributes anything or where players invest beyond the threshold value e^*. For the star the result dictates that the center player must contribute strictly more than every peripheral player, as long as at least one of them makes a positive contribution as well. However, fair play also allows the case where nobody contributes anything.

The prediction is of lesser use in the circle, where it is clearly fair if all players contribute the same amount. But, if the sequence of players is $i - j - k - l$, fair play also allows the profile $e_i > e_j = e_l > e_k$ because this implies $b_i > b_j = b_l > b_k$ such that payoffs are equalized.

3.3 Marginal fairness concerns

Another possible fairness norm is one, where all players care about payoff differentials but only to a marginal extent. This means that α_i and β_i in utility function (2) are both positive but small.

3.3.1 What is marginal?

Players with marginal fairness concerns put most weight on their own payoffs. As a consequence, their best response is often indistinguishable from selfish play. Specifically, for total contributions significantly different from the payoff-maximizing level e^*, where - due to the concavity of $b(\cdot)$ - the monetary consequences of an additional contribution are significant, a player will behave as if she was selfish. Only for contributions close to e^*, where a marginal deviation is not costly, does player behavior display fairness concerns. Expanding on the Bramoullé and Kranton (2007) best response function of a selfish player, the following definition formalizes this:

Definition 1. *For any g and profile $e_{-i} = \{e_1, e_2, ..., e_{i-1}, e_{i+1}, ..., e_n\}$ the optimal reaction of player i satisfies:*

1. $e_i = e_i^* - \sum_{j \in N_i(g)} e_j$ *if and only if* $\sum_{j \in N_i(g)} e_j \leq e_i^*$

2. $e_i = 0$ *if and only if* $\sum_{j \in N_i(g)} e_j \geq e_i^*$

where $e_i^* \in \left[e^* - \underline{\eta}\, , \, e^* + \overline{\eta} \right]$ *and* $\underline{\eta}, \overline{\eta} > 0$ *but "small".*

Moreover, if $\pi_i \geq (\leq)\ \pi_j$ *for all* $j \in N_i(g)$ *with at least one inequality being strict, then* $e_i^* > (<)\ e^*$*.*

Just like a selfish player, a marginally fairness concerned player invests up to (but no more than) her personal maximum contribution level of e_i^*. The difference is that this value depends on her strength of inequity aversion as well as on neighbors' contributions. Depending on the combination of the two, e_i^* may lie above e^* or below. Nevertheless, the personal maximum is close to the selfish benchmark.

Specifically, there are some upper and lower bounds on e_i^* that are common to all players in the same game. The bounds are determined by the situations in the game that trigger the most extreme response of fairness concerns . At the one extreme, the player with the strongest feeling of guilt is put in a position, where an additional contribution to the public good has the largest positive impact on her neighbors' payoffs (relative to her own gains or losses). This player defines $e^* + \overline{\eta}$. At the other extreme, the most envious player is in a position, where an incremental contribution reduction hurts her neighbors the most (on top of her own cost savings) . That player defines $e^* - \underline{\eta}$. Thus, the bounds are in fact not defined by the most unfair situations in a game, in terms of payoff differences, but rather by the situations in which a player's incremental contribution has the the largest impact on others.

More concretely, $\overline{\eta}$ is defined by the following two-step procedure: for a given structure g and the largest guilt parameter $\bar{\beta} = \max_{i \in N}\{\beta_i\}$ (which is still sufficiently small so that utility function (2) inherits the functional properties of $b(\cdot)$, i.e. $U(e = 0) > c$ and $U''(\cdot) < 0$ in all its arguments on the differentiable domain of the function), choose e'_i and $e'_{-i} = \{e_1, e_2, ..., e_{i-1}, e_{i+1}, ..., e_n\}$ for all $i \epsilon N$ such that:

$$e' = (e'_i\, , \, e'_{-i}) = \arg\min_e \left\{ \sum_{j \in N_i(g)} [\frac{\partial b_i}{\partial e_i} - \frac{\partial b_j}{\partial e_i}] \right\} \qquad (3)$$

under the constraints that $\pi_i(e') > \pi_j(e')$ for all $j \in N_i(g)$ and that the first-order condition of utility function (2) is satisfied for $e'_i + \sum_{j \in N_i(g)} e'_j$, i.e.:

$$\left.\frac{\partial U_i}{\partial e_i}\right|_{e'} = \frac{\partial b_i}{\partial e_i} - c - \bar{\beta} \sum_{j \in N_i(g)} \left.\left[\frac{\partial b_i}{\partial e_i} - c - \frac{\partial b_j}{\partial e_i} \right]\right|_{e'} = 0.$$

Then, the upper bound is defined by the position in g with

$$\overline{\eta} \equiv \max_{i \epsilon N} \left\{ e'_i + \sum_{j \in N_i(g)} e'_j \right\} - e^*.$$

In the star structure of Figure 1, for example, the center player has the largest positive impact on her neighbors' payoffs, when $e'_c = 0$ and $e'_{1_p} = e'_{2_p} = e'_{3_p} > 0$ is such that e'_p minimizes $\left\{ \frac{\partial b_c}{\partial e_c} - \frac{\partial b_p}{\partial e_c} \right\}$, while $3e'_p$ satisfies the center's first-order condition. Periphery player 1, on the other hand, has the biggest impact, when $e'_{2_p} = e'_{3_p} = 0$, but for any $e'_c > e'_{1_p} \geq 0$ that satisfies player 1's first-order condition, because reducing e'_{2_p}, e'_{3_p} increases $\frac{\partial b_c}{\partial e_1}$, and if $e'_{2_p} = e'_{3_p} = 0$ then $\frac{\partial b_1}{\partial e_1} = \frac{\partial b_c}{\partial e_1}$. Hence, (3) becomes $\arg\min \{0\}$, and $\overline{\eta}$ is therefore defined by putting the most guilty player in the star center position.

Similar, for a given g and the largest envy parameter $\bar{\alpha} = \max_{i \in N}\{\alpha_i\}$, a conservative lower bound $\underline{\eta}$ is given by:

$$e'' = (e_i'', e''_{-i}) = \arg\max_{e} \left\{ \sum_{j \in N_i(g)} [\frac{\partial b_j}{\partial e_i} - \frac{\partial b_i}{\partial e_i}] \right\},$$

under the constraints that $\pi_i(e'') < \pi_j(e'')$ for all $j \in N_i(g)$ and that the first-order condition of (2) is satisfied for $e''_i + \sum_{j \in N_i(g)} e''_j$, i.e.:

$$\left.\frac{\partial U_i}{\partial e_i}\right|_{e''} = \frac{\partial b_i}{\partial e_i} - c - \bar{\alpha} \sum_{j \in N_i(g)} \left[\frac{\partial b_j}{\partial e_i} - \frac{\partial b_i}{\partial e_i} + c \right] \Bigg|_{e''} = 0.$$

The lower bound is then determined by position $i \epsilon N$ with

$$\underline{\eta} \equiv e^* - \min_{i \epsilon N} \left\{ e''_i + \sum_{j \in N_i(g)} e''_j \right\}.$$

A contribution reduction by the star center, for example, is most detrimental for her neighbors, when $e'_c > e'_p > 0$ is such that it maximizes $\left\{ \frac{\partial b_p}{\partial e_c} - \frac{\partial b_c}{\partial e_c} \right\}$, while $e'_c + 3e'_p$ satisfies the center's first-order condition. For periphery player 1, this is achieved, when $e'_{2_p} = e'_{3_p} = 0$, but for any $e'_{1_p} > e'_c \geq 0$ as long as player 1's first-order condition is satisfied. Thus, $\underline{\eta}$ is defined by the most envious player in the center position.

A player's maximum deviation from selfish play, $e^*_i - e^*$, comes from the interval spanned by these bounds $[-\underline{\eta}, +\overline{\eta}]$. The maximum deviation can also be given a monetary interpretation, as the payoff difference $b(e^*_i) - b(e^*) - c(e^*_i - e^*)$ measures how much a player is maximally willing to sacrifice in order to achieve a more fair outcome.

3.3.2 Equilibrium predictions

Despite the incremental differences to the definition of a selfish best response, the following results show that marginal fairness concerns lead to significantly refined, and occasionally surprising, equilibrium predictions. In fact, in a local public goods game, marginal fairness selects among the set of equilibria when selfish play allows multiple equilibria for a given interaction structure. We begin with our predictions for the interaction structures underlying most prior public goods experiments, namely the dyad and the complete interaction structure. The following result shows that even marginal fairness concerns are sufficient to produce an extremely fair outcome:

Proposition 2. *Consider a perfectly symmetric and fully transitive interaction structure with $N \geq 2$. Suppose $\alpha_i, \beta_i > 0$ for all $i \in N$ but with all values being small. An equilibrium contribution profile satisfies $e_i = e_j \in \big[\frac{e^*-\underline{\eta}}{N}, \frac{e^*+\overline{\eta}}{N}\big]$ for all $i, j \in N$.*

Proof. Our first aim is to prove that in equilibrium it needs to hold that $e_i = e_j$ for all $i, j \in N$. We then proceed to show that the total contribution of all players must be close to e^*.

To prove the first, note that for any two players, i and j, in the perfectly symmetric and fully transitive interaction structure it holds that $b_i = b_j$. Hence, payoff differences are solely due to differences in contributions. Now suppose a contribution profile exists where two players, i and j, invest $e_i > e_j \geq 0$. Let i be the player with the highest contribution in the game and j the one with the lowest contribution. As player i is exploited by at least one other player and there is no other player with a lower payoff, it follows from Definition 1 that $e_i^* < e^*$. At the same time, for player j, $e_j^* > e^*$. This, however, leads to a clash with the equilibrium conditions because an equilibrium simultaneously requires:

$$e_i + \sum_{j \in N_k(g)} e_k = \sum_{k \in N} e_k = e_i^* < e^* \quad \text{and} \quad e_j + \sum_{k \in N_j(g)} e_k = \sum_{k \in N} e_k \geq e_j^* > e^*$$

Thus, in equilibrium it must be that $e_i = e_j$ for all $i, j \in N$.

It remains to be shown that the total contribution of all players is close to e^*. Obviously, it cannot be that $e_i = e_j = 0$ for all $i, j \in N$, which follows from part (2) of Definition 1. Moreover, it must be that $e_i^* = e_j^*$, which follows from part (1) together with the fact that $e_i = e_j$. Hence, it must be $e_i = \frac{e_i^*}{N} \in \big[\frac{e^*-\underline{\eta}}{N}, \frac{e^*+\overline{\eta}}{N}\big]$ for all $i \in N$. □

The intuition behind Proposition 2 is the following: in a perfectly symmetric and fully transitive structure, all players have access to the same amount of the public good. Payoff differences are solely based on differences in contributions. When players are marginally fairness concerned, they will each choose a contribution level ensuring the total is close to the payoff-maximum level of e^*. However, on this domain of the benefit function, $b(\cdot)$, an incremental change in a player's contribution is hardly payoff relevant to her. Hence, the player feels the need to equalize payoff differences. Players who contribute less than the other players will add a unit, while those over-investing will cut back. Fairness concerns are absent only when all players invest exactly the same. The result shows that equilibria can be maintained where players deviate from the selfish equilibrium, e^*. The reason is that, given all other players contribute e_i^*/N, any unilateral deviation from $e_i = e_i^*/N$ would be prevented by a focal player's fairness concerns. As long as the material benefit from deviating is smaller than the relative cost of more inequality, she will maintain the equilibrium level.

Some aspects of this result are worth pointing out. Since all players are predicted to make exactly the same contributions, marginal fairness concerns are sufficient to produce an outcome that looks remarkably fair. In fact, the predictions differ from pure fair play only in that total contributions are close to the benchmark value of selfish play, e^*. This is noteworthy because the vast majority of experiments are conducted in the very same setting of the complete interaction structure investigated here. However, the result suggests that the symmetry of the setting itself makes it difficult to distinguish between significant or merely marginal fairness concerns because a fair outcome is prevalent even under marginal fairness concerns.

We continue with our predictions for the star:

Proposition 3. *Consider a star with $N > 2$. Suppose $\alpha_i, \beta_i > 0$ for all $i \in N$, but with all values being small. In equilibrium it must be that $e_c = 0$ for the center player i_c and $e_p \in [e^* - \underline{\eta}\,, e^*)$ for all peripheral players j_p.*

Proof. We prove the claim by stepwise exclusion of out-of-equilibrium profiles. Clearly, $e_i = e_j = 0$ for all players $i, j \in N$ does not establish an equilibrium, which follows from part (2) of Definition 1. In addition, it must be that $e_j > 0$ for at least one peripheral player, because suppose otherwise and $e_j = 0$ for all j and $e_i > 0$. Then, following from part (1) of Definition 1, $e_i = e_i^*$, where clearly $e_i^* < e^*$ because the center player feels envy as she feels exploited by the peripheral players. However, at the same time, $e_j^* > e^*$ for all peripheral players. It follows that $e_c^* < e_p^*$, which is incompatible with part (2) of Definition 1 requiring that $e_i \geq e_j^*$.

Thus, $e_j > 0$ must hold for at least one peripheral player. We continue showing that this implies $e_i = 0$ in equilibrium. Suppose to the contrary that $e_i > 0$. Two different cases may arise:

1. There is only one j with $e_j > 0$. In this case, by part (1) of Definition 1 the contribution received by j satisfies $e_j + e_i = e_j^*$, whereas by part (2) for any k with $e_k = 0$, it is $e_i \geq e_k^*$. Hence, because the maximum contribution values of all players are sufficiently close, we additionally obtain the following chain of inequalities:

$$\overline{\eta} + \underline{\eta} \geq e_j^* - e_k^* = e_i + e_j - e_k^* \geq e_i + e_j - e_i = e_j$$

   This, however, results in a contradiction to $\overline{\eta} + \underline{\eta}$ being small because the contribution received by the center player is no larger than $\overline{\eta} + \underline{\eta}$, which implies that $e_i = e_i^* - e_j \geq (e^* - \underline{\eta}) - (\overline{\eta} + \underline{\eta}) = e^* - \overline{\eta} - 2\underline{\eta}$. Hence, for $\overline{\eta}$ and $\underline{\eta}$ being small, we obtain $\pi_i < \pi_j$ for all $j \neq i$ and, in turn, $e_i^* = e_i + e_j < e^*$. This results in a contradiction to $e_j^* = e_i + e_j > e^*$.

2. There is more than one j with $e_j > 0$. Because the maximum contribution values of all players are sufficiently close, it must hold that $\overline{\eta} + \underline{\eta} \geq e_i^* - e_j^* = \sum_{l \neq i,j} e_l$ for center player i and any peripheral player j with $e_j > 0$ (and similar for any other peripheral player k with $e_k > 0$). This means that the total contribution received by center player i is $\sum_{j \neq i} e_j \leq \sum_{l \neq i,j} e_l + \sum_{m \neq i,k} e_m \leq 2(\overline{\eta} + \underline{\eta})$. Hence, $e_i = e_i^* - \sum_{j \neq i} e_j \geq (e^* - \underline{\eta}) - 2(\overline{\eta} + \underline{\eta}) = e^* - 2\overline{\eta} - 3\underline{\eta}$. Then, for sufficiently small $\overline{\eta}$ and $\underline{\eta}$, it is $\pi_i < \pi_j$ for all $j \neq i$ with $e_j > 0$, and $e_i^* = e_i + \sum_{j \neq i} e_j < e^*$, which cannot be aligned with $e_j^* = e_i + e_j > e^*$, which needs to be true if $\pi_i < \pi_j$ holds.

Hence, in equilibrium it must be that $e_i = 0$ and therefore, by part (2) of Definition 1, also $e_j > 0$ for all peripheral players. In fact, the definition implies that $e_j \in [e^* - \underline{\eta}\,, e^*)$. □

Some aspects of this result noteworthy: again, just as for the complete interaction structure, marginal fairness concerns allow for a sharp prediction about equilibrium play. This equilibrium produces a highly unequal payoff distribution, which is reminiscent of the familiar predictions for selfish behavior in public goods games: the star center player entirely free rides on the peripheral players' contributions. Note also that this equilibrium is consistent with the findings from prior experimental studies that behavior typically does not show fairness concerns when players influence each other in a heterogeneous way.

The intuition behind this can be found in Definition 1: suppose the center player would make all the contributions and the peripheral players free ride. The center player would not contribute up to e^* because she rightfully perceives the situation as unfair. Hence, from the peripheral players' perspectives, there are still some private gains to be made and, combined with their own fairness concerns, they contribute to the public good themselves. Hence, a center-sponsored public good (which is actually an equilibrium under selfish play) is unstable when players are marginally fairness concerned. Building on this, suppose a peripheral player would at least make a small contribution to the public good. The center player could then afford to cut back on her own contributions. This, in turn, implies additional contribution incentives in the periphery. Continuing along this line, it follows that the only equilibrium profile is one where the center player free rides entirely.

We continue with the circle, which is another symmetric interaction structure, but – unlike the complete interaction structure – the interactions on the circle have gaps (intransitivities). Since it is difficult to make clear predictions for a circle of arbitrary size, we confine ourselves to the case of four players, which is also the number of players in our experiment.[9]

[9] Even though we were not able to develop an exhaustive characterization similar to Proposition 4 for larger circles of $N > 4$, the two classes of equilibria characterized in the proposition are also equilibria in those cases.

Proposition 4. *Consider a circle with $N = 4$. Suppose $\alpha_i, \beta_i > 0$ for all $i \in N$, but with all values being small. In equilibrium, either:*

- *$e_i \in \left[\frac{e^*}{3} - \frac{2\overline{\eta}+3\underline{\eta}}{3} \,;\, \frac{e^*}{3} + \frac{3\overline{\eta}+2\underline{\eta}}{3}\right]$ for all $i \in N$, or*
- *for any pair of neighbors i, j: $e_i \in [e^* - \underline{\eta}\,,\, e^*)$ and $e_j = 0$.*

Proof. Let us fix the sequence of players in the following order: $i - j - k - l$. First, suppose that $e_i > 0$ for all $i \in N$. According to part (1) of Definition 1, it holds that $e_l + e_i + e_j = e_i^* \in [e^* - \underline{\eta}\,,\, e^* + \overline{\eta}]$. Thus, a lower bound for player i's contribution is:

$$e_l + e_i + e_j \geq e^* - \underline{\eta} \tag{4}$$

for given e_j and e_l, and similarly, for players j, k, l. Moreover, in a circle of four it follows for players j and k that their maximum contribution values are sufficiently close. Hence:

$$e_k^* - e_j^* = e_j + e_k + e_l - (e_i + e_j + e_k) = e_l - e_i \leq \overline{\eta} + \underline{\eta} \tag{5}$$

and similarly for $e_j^* - e_k^*$ and all other pairs ij, kl, and li. Hence, suppose e_i is the smallest contribution value. Then (4) and (5) combined result in the following lower bound:

$$e_i \geq e^* - \underline{\eta} - e_j - e_l \geq e^* - \underline{\eta} - 2(e_i + \overline{\eta} + \underline{\eta}) \quad \Leftrightarrow \quad 3e_i \geq e^* - 2\overline{\eta} - 3\underline{\eta}$$

This is equivalent to the lower bound stated in the proposition.

Similarly, to find an upper bound for the highest value e_i it holds:

$$e_i \leq e^* + \underline{\eta} - e_j - e_l \leq e^* + \overline{\eta} - 2(e_i - \overline{\eta} - \underline{\eta}) \quad \Leftrightarrow \quad 3e_i \leq e^* + 3\overline{\eta} + 2\underline{\eta}$$

which is equivalent to the upper bound stated in the proposition.

Next, suppose that in equilibrium $e_i = 0$ for at least one player in the circle $i - j - k - l$. It follows that by part (2) of Definition 1 player i's neighbors j and l must invest enough such that $e_j + e_l \geq e_i^*$. Moreover, it must be that $e_j > 0$ and $e_l > 0$ in equilibrium because suppose to the contrary that $e_j = 0$ (or $e_l = 0$ or both are equal to zero). Then $e_k > 0$ since otherwise $e_i + e_j + e_k = 0$. In fact, we would require simultaneously that $e_k \geq e_i^* \geq e^* - \underline{\eta}$ and $e_l \geq e_j^* \geq e^* - \underline{\eta}$. This immediately leads to a contradiction to part (1) of Definition 1 because it implies for player k: $e_j + e_k + e_l \geq 2(e^* - \underline{\eta}) > e_k^*$ and the same for l.

Thus, if $e_i = 0$ in equilibrium, $e_j > 0$ and $e_l > 0$ must hold. But this also implies $e_k = 0$ because suppose to the contrary that $e_k > 0$. As the maximum contribution values are sufficiently close together, it holds that $\overline{\eta} + \underline{\eta} \geq e_k^* - e_j^* = e_l$ and similarly $\overline{\eta} + \underline{\eta} \geq e_k^* - e_l^* = e_j$. This implies, however, that the total contributions received by player i are no larger than $2(\overline{\eta} + \underline{\eta})$. Hence, for sufficiently small $\overline{\eta}$ and $\underline{\eta}$ it is $e_j + e_l < e_i^*$. A contradiction to $e_i = 0$ being played in equilibrium. Hence, $e_k = 0$ must hold. Moreover, the upper and lower bounds for e_j, e_l follow from Definition 1. □

The predictions for the circle are a mixture of those for the star and the complete interaction structure. On the one hand, a range of almost equal contributions can be supported in equilibrium, which also comprises the profile of identical contributions that we found for the complete interaction structure. This seems natural given the symmetry of the circle. On the other hand, a highly unequal contribution profile may emerge, which is reminiscent of the star. This is surprising because it suggests that a mere intransitivity in the interaction structure is sufficient for highly unequal outcomes in symmetric settings when people are only marginally fairness concerned.

The intuition is the following: unlike in a fully transitive structure, the players in a circle do not interact with every other player. In particular, the players who maintain high contributions in the unequal equilibrium do not interact with each other. Even though they might feel exploited by their direct neighbors, they prefer to maintain a high level of contributions for the sake of their own payoffs. The free riders therefore receive a total contribution far beyond their personal maximum level and, being only marginally fairness concerned, see no reason to bear the extra cost of a less unequal outcome.

3.4 Power- and need-based norms

We have so far developed predictions for equilibrium play under various intensities of fairness concerns but omitted considering the possibility that the distributive fairness norm shared by the subjects in our experiment might also depend on the interaction structure itself. In particular, subjects might share the view that those in a more central position deserve a higher payoff because of their more powerful position. Under the most extreme form of such a power-based norm, a player feels envy only if another player in a less central position earns a higher payoff and guilt only when a more central player earns less. Alternatively, the very same position might become an obligation when under a need-based norm the center player is expected to help those in inferior positions.

Here we consider such position-dependent distributive norms and assume that need and power are derived from the most obvious features of the interaction structure: the player's degree centrality in an interaction structure. We expect for the star and a power-based norm that the center player feels more envy and less guilt than the peripheral players, whereas the opposite holds true for a need-based norm. In the complete interaction structure, the dyad, and the circle, on the other hand, there is no degree asymmetry, and, hence, we do not expect position-based norms to play a role there. The following formal results focus on strong fairness concerns where players try to avoid, in the first instance, feelings of envy or guilt.[10] Moreover, we assume that under a power-based norm the feeling of guilt is entirely absent for the star center

[10] The theory of marginal fairness concerns introduced above allows players to differ in their concerns to some extent. Hence, the predictions of this theory also apply to positioned-based fairness concerns as long as they are small.

player, whereas peripheral players feel no envy. Likewise, we assume that the center player does not feel envy and the peripheral player does not feel guilt under a need-based norm. For players in the complete interaction structure or in the circle, however, the pairwise comparisons work in the same way as in utility function (2). Formally, we rewrite (2) under a power-based norm in the following way:[11]

$$U_i = \pi_i - \sum_{\substack{j \in \overline{N}_i(g) \cap \\ |N_i(g)| > |N_j(g)|}} \alpha_i \pi_j - \sum_{\substack{j \in \underline{N}_i(g) \cap \\ |N_i(g)| \leq |N_j(g)|}} \beta_i \pi_i \qquad (6)$$

where $|N_i(g)|$ denotes the number of neighbors of player i and $\overline{N}_i(g)$ ($\underline{N}_i(g)$) is the subset of neighbors who earn weakly more (strictly less) than player i. Under a need-based norm, we replace the set of neighbors in the first sum by $j \in \overline{N}_i(g) \cap \big(|N_i(g)| < |N_j(g)|\big)$ and by $j \in \underline{N}_i(g) \cap \big(|N_i(g)| \geq |N_j(g)|\big)$ in the second sum, respectively.

Turning to the equilibrium predictions, they are obviously indistinguishable from those of the pure fairness theory in case of the symmetric interaction structures: the dyad, the complete interaction structure, and the circle. The following two results summarize our predictions for the star:

Proposition 5. *Consider a star with $N > 2$. Suppose a power-based norm and suppose that α_i and β_i are both large. In equilibrium, it must be that $e_i = 0$ for the center player i and $e_j = e^*$ for all peripheral players j.*

Proof. Obviously, any contribution profile where $\pi_i \leq \pi_j$ for the center player i and at least one peripheral player j, cannot establish an equilibrium because both i and j would want to avoid their respective feelings of envy or guilt. Thus, $\pi_i > \pi_j$ for center player i and all peripheral players j. Player i's and j's first-order conditions of (6) can then be written as:

$$\frac{dU_i}{de_i} = \frac{db_i}{de_i} - c \leq 0$$

$$\frac{dU_j}{de_j} = \frac{db_j}{de_j} - c \leq 0$$

In fact, these two lines are exactly the same conditions of the Bramoullé and Kranton (2007) theory of purely selfish play. The authors show that, together with $\pi_i > \pi_j$, they can only be satisfied simultaneously if $e_i = 0$ and $e_j = e^*$. □

[11] We depart from the original model by an additional detail, namely that a player considers the entire payoff of a richer but less central, or a poorer but more central, neighbor as a negative. In the original model, a player considers only the difference to his or her own payoff as a negative. Both specifications reproduce the same notion that fairness concerned players aim to reduce payoff differences. The reason for this alternative specification is a technical finesse as it avoids the problem of having a kink in the marginal utility function at the point where two players earn exactly the same profit.

As expected, the power-based norm promotes a highly unequal payoff distribution. Our prediction for a need-based norm is the following:

Proposition 6. *Consider a star with $N > 2$. Suppose a need-based norm and that α_i and β_i are both large. In equilibrium, it must be $e_i = e^*$ for the center player i and $e_j = 0$ for all peripheral players j.*

The proof is analogous to the proof of Proposition 5. Obviously, the power-based norm is able to explain the findings of prior experiments, namely that fairness concerned behavior was switched off as soon as subjects played cooperation or bargaining games on heterogeneous interaction structures or under threat of (one-sided) competition (Roth, Prasnikar, Fujiwara, and Zamir, 1991; Harrison and Hirshleifer, 1989; Plott, 1982). It remains to be seen how well this norm explains experimental behavior in our local public goods games.

3.5 Summary of predictions

Table 1 summarizes the equilibrium predictions for the different theories and interaction structures. Some remarks may be helpful on the Bramoullé and Kranton (2007) theory of pure selfishness, which we have not fully described yet. According to this theory, any combination of player contributions satisfying $\sum_{i \in N} e_i = e^*$ establishes an equilibrium on the dyad and the complete interaction structure. In order to refine their predictions, the authors have also looked at stable equilibria based on the Nash tâtonnement process (Fudenberg and Tirole, 1991). However, the concept is not applicable to these two interaction structures since no equilibria are stable at the same time. Hence, we use the predictions for these two interaction structures to discriminate between purely selfish and marginal fairness concerned behavior because, according to the latter, players coordinate on one particular equilibrium of the many selfish equilibria. In the star and circle, on the other hand, stability selects one of two possible selfish equilibria, which are marked with an exclamation mark (!) in Table 1. The stable equilibrium in the star is the same as that selected by marginal fairness theory.

Comparing the predictions for pure and marginal fairness concerns, note that the former theory requires the star center player to invest more than any peripheral player as a compensation for their inferior positions. Thus, unlike in the complete interaction structure and the dyad, the intensity of fairness concerns plays a crucial role for the predicted contributions in the star. In this sense, the star provides the ultimate test bed for whether players are "truly" or just marginally fairness concerned.

(Table 1 about here)

4 Experimental setup and procedure

We ran a computerized experiment, programmed in z-tree 3.0 (Fischbacher, 2007) and administered in the Experimental Laboratory for Sociology and Economics (ELSE) at Utrecht University in The Netherlands. Using the ORSEE recruitment system (Greiner, 2004), we invited more than 1,000 students across all disciplines to participate in the experiment. In total, we conducted eight experimental sessions of approximately one and a half hours each. A total of 120 subjects participated (15 students per session on average).[12] In each session we administered the seven interaction structures (henceforth treatments) as shown in Figure 1. The order of the treatments was balanced across the sessions. The instructions, as shown in the Appendix, were handed out before the start of the experiment.

Within each treatment, subjects played the local public goods game on a given interaction structure in 5 repetitions (rounds). Each treatment consisted of one trial round and four payment-relevant rounds (altogether 35 rounds of which 28 were relevant for subjects' earnings). At the beginning of each round, subjects were randomly allocated to a group together with either one (in the dyad) or three other subjects (in all other structures). Thus, every subject played 35 rounds in 35 different groups.[13] This resulted in 3,360 contribution decisions that were payoff relevant (120 subjects times seven treatments times four rounds). For the four structures in Table 1 we obtained 1,920 contribution observations (120 subjects times four treatments times four rounds). For each unique position in a structure (e.g., the star center player) there are 120 observations (120 subjects divided by four players in a group times four rounds).

Each round had the same structure and lasted between 30 and 90 seconds, each ending at an unknown and random moment during this time interval. Starting from a situation with no contributions, subjects indicated simultaneously on their computer terminals how much they wished to invest. Full information about the contributions of all other subjects was continuously provided. Moreover, Furthermore, resulting payoffs of all participants could continuously be observed on the screen and were easily identifiable by the size of the bubble around own and other subjects' decision nodes.[14] At the end of each round, subjects were informed about the number of points earned, based on their own and their neighbors' contributions. In other words, final earnings depended solely on the situation at the end of a round.

Experience with previous experiments on behavior in complex interaction structures suggests individuals find it difficult to coordinate their behavior (see, e.g., Rosenkranz and Weitzel, 2012, Goeree, Riedl, and Ule, 2009, Falk and Kosfeld, 2012). Subjects appear to need time to understand the game and coordinate their actions. One way to facilitate coordination is to simplify subjects' decisions to dichotomous choices (e.g., invest vs. do not invest). Given our focus on inner solutions to, and marginal deviations from, an optimal contribution decision, this is not feasible in our case. We therefore facilitate coordination by giving subjects ample time to coordinate, first, by starting each

[12] As the structures needed groups of four, we ran sessions with either 12, 16, or 20 students. The average age of the participants was 22 years; 67 percent were female and 72 percent had Dutch nationality.

[13] We ensured that no group was randomly drawn twice.

[14] See the screenshot in the instructions in the Appendix.

treatment with a pure trial round (not payoff relevant) of 90 seconds. Second, we commence each subsequent round in a treatment with a coordination period of 30 seconds, where actions do not have any payoff implications. After this coordination period, the round was randomly stopped within one minute to determine the payoff relevant contribution profile. This experimental design emulates a sequence of simultaneous move games with a stochastic end stage, where only the last stage behavior is payoff relevant. In such a game, any equilibrium of the stage game can be implemented in the stage game of our experiment. Actions in the payoff irrelevant coordination period can be interpreted as cheap talk, which, as equilibria in our game are not Pareto ranked, should neither select any equilibrium nor lead to any new equilibria.

The payoffs at the end of a round were determined in line with equation (1) as follows:

$$\pi_i = (e_i + \sum_{j \in N_i(g)} e_j)(29 - (e_i + \sum_{j \in N_i(g)} e_j)) - 5e_i. \tag{7}$$

The maximum contribution level e^* is a joint contribution of 12 points, which subjects had to find out themselves. All points earned were converted into euro at an exchange rate of 400:1. On average, subjects earned 11.82 euro .

5 Results

5.1 Data and measures

To test our equilibrium predictions of Table 1, we focus on the smallest homogeneous unit: unique node positions in each of the four interaction structures. In symmetric structures, all node positions are the same, but in the star, for example, there are two different node positions: the center and the periphery.

We study the recorded data along three dimensions. First, we compute the frequencies of specific contribution levels. For each node position, the x-axis in Figure 2 shows the amount a subject could have invested at the end of a round and the y-axis shows how often this contribution level was chosen (as a percentage of all decisions in that node position). Note that contributions were only possible in integers. Therefore, each bar in Figure 2 represents exactly one possible contribution level. Besides, note that the star center has only 120 observations (as explained in Section 4), while the star periphery has three times as many nodes and thus 360 observations.

(Figure 2 about here)

Second, we compare subjects' contribution decisions to the best response behavior of purely selfish players. Fairness concerns can lead to deviations from this behavior, which we compute as follows:

$$D_i(BR_i) = \begin{cases} e_i & \text{if } \sum_{j \in N_i(g)} e_j \geq 12 \\ (\sum_{j \in N_i(g)} e_j + e_i) - 12 & \text{if } \sum_{j \in N_i(g)} e_j < 12 \end{cases}$$

As $e^* = 12$, any $e_i > 0$ is a positive deviation from best response if the neighborhood jointly invests more than 12. With a neighborhood jointly investing less than 12, any e_i that fills the gap to 12 too much (too little) is a positive (negative) deviation from best response. Figure 3 reports the frequency distributions of best response deviations per node position.

(Figure 3 about here)

Third, for every node we compute the difference to the contribution of each of its neighbors. This measure is useful to compare observed decisions with behavior under pure fairness. Given the predictions for the model with pure fairness concerns, the deviation from the contribution of neighbors should be zero for the dyad and the complete interaction structure. Specifically, we compute:

$$D_i(e_j) = e_i - e_j \quad \forall j \in N_i(g)$$

Figure 4 shows the distribution of $D_i(e_j)$ per node position. Note that the number of observations is now multiplied by the number of neighbors per position. For each of the 120 decisions at the star center, for example, we have three differences in contributions, one for each of the three neighbors (360 observations). In the complete interaction structure, this applies to all four nodes ($360 \times 4 = 1440$ observations in total).

(Figure 4 about here)

As Figures 2, 3, and 4 (and the underlying data in Tables 4, 5 and 6, in the Appendix) show, contribution behavior is remarkably organized and well explained by our static theory predictions despite the dynamics of the game and the coordination problems that subjects might have run into. Specifically, behavior strongly depends on the characteristics of the node position. In general, the equitable outcomes seem to be typical for the complete interaction structure and the dyad. Additionally, for the star, subjects in the center position clearly invest less while those in the periphery invest more. This dependency can also be observed in the other three interaction structures we administered in the experiment (see Appendix). In the following, we investigate these observations in more detail and confront them with the predictions from the theories derived in Section 3.

5.2 Evidence for pure selfishness

For the two symmetric structures, the dyad and the complete interaction structure, the theory of purely selfish play predicts multiple equilibria, none of which is stable. Hence, any distribution of individual contributions would be conceivable. For the circle and the star, one stable equilibrium profile is predicted: in the circle, every second subject free rides while the others choose $e^* = 12$. In the star all peripheral players invest $e^* = 12$ while the center player free rides. Moreover, all contributions in all interaction structures should be best responses to each other. Furthermore, we expect 100% of the star center players to deviate by-12 from their neighbors, whereas for the circle we expect a two-point distribution where half of the players deviates by-12 from their neighbors while, accordingly, the other half deviates by $+12$.

The predictions of selfish play are most strongly supported in the star. We find only mixed evidence for the dyad, the complete interaction structure, and the circle structure. Figure 3 and Table 5 show that subjects in most interaction structures either best respond to their neighbors or deviate by only a single unit (58.13% in the dyad, 53.75% in the circle, 79.17% in the star center, and 67.5% in the star periphery). The weakest results along this dimension are obtained for the complete interaction structure, where only 47.71% are close to their best response. As predicted by selfishness theory and illustrated in Figure 2 and Table 4, subjects indeed coordinate on the periphery-sponsored star: 62.50% of the star center players invest exactly 0 units and another 15.0% invest 1 unit. At the same time, in the star periphery 39.72% of our subjects invest 12 units and another 13.6% invest 11 units.

Although our findings for the dyad and the complete interaction structure are largely also consistent with the predictions of selfish play, the most striking observation is that subjects seem to coordinate on one of the many possible equilibria, namely the one inducing an equal payoff distribution: Figure 2 and Table 4 show that in the dyad 42.5% invest exactly $\frac{e^*}{2} = 6$ units while another 28.75% invest either 5 or 7 units (i.e., 14.58% and 14.17%, respectively). Thus, 71.25% of subjects choose either equal or nearly equal contributions. In the complete interaction structure, these shares are clearly lower. However, a considerable 26.67% of our subjects invest $\frac{e^*}{4} = 3$ units, and 33.96% invest either 4 or 2 units (i.e., 22.5% and 11.46%, respectively). An explanation of these focal equilibria is outside the scope of the selfishness theory but can be found in the fairness literature.

Our results for the circle do not provide much support for selfish play: Table 4 shows that only 19.17% and 6.88% of our subjects play either 0 and 12, respectively, or deviate by no more than a single unit from these values. Instead, a substantial share of subjects choose contribution levels between 3 and 5 units, namely 37.92%, indicating that some groups coordinate on an equal payoff distribution. This is also supported by the three-mode distribution we find for the circle in Figure 4 (Table 6). Overall, therefore, selfishness theory performs well for the star but proves problematic in predicting behavior in symmetric interaction structures.

5.3 Evidence for pure fairness

According to pure fairness theory, contribution profiles should be characterized by $e_i = e_j \geq 0$ in the dyad, the complete interaction structure, and the circle, whereby in the latter a fair outcome can also be achieved when $e_i > e_j = e_l > e_k$. In the star we should observe $e_c \geq e_p$. This implies that we expect $D_i(e_j) = 0$ for the dyad and the complete interaction structure while $D_c(e_p) > 0$ and $D_p(e_c) < 0$ in the star. In the circle we either expect that all contributions are equal or that the deviation from neighboring contributions follows a bimodal distribution, with one mode being at point x, where $x \epsilon (0, 12)$, and the other mode at $-x$.

In the complete interaction structure, as seen in Figure 4 and Table 6, a significant share of players do not deviate from their neighbors by more than one unit (51.11%). In the dyad it is an even more substantial share (70.83%). In the circle, 14.58% of the contributions are identical to a neighboring contribution and 31.04% deviate at most by one unit. Other subjects coordinated on highly unequal contribution levels in the circle: Figure 4 shows that 13.54% of the decisions deviate by either +12 or -12 units from their neighbors. However, such large deviations cannot be explained by pure fairness concerns but rather reflect extremely unfair, specialized equilibria. The clearest evidence against fairness theory comes, however, from the star: in 94.7% of all cases, we find $D_c(e_p) \leq 0$ and $D_p(e_c) \geq 0$. Hence, our findings provide support for pure fairness theory only in fully symmetric and transitive structures (dyad and complete). But even there, pure fairness theory cannot explain why subjects are apparently unwilling to sacrifice a great deal for a fair payoff distribution. As Figure 3 shows, most rounds end with a total contribution of close to 12, although any other profile with equal contributions would be equally fair but yield lower monetary payoffs.

5.4 Evidence for marginal fairness concerns

Marginal fairness theory predicts that individual contributions in the dyad and the complete interaction structure are close, but not necessarily identical, to $e_i = e_j \approx \frac{e^*}{N}$. In the circle, contribution profiles are predicted to be either $e_i \approx \frac{e^*}{3}$ for all players or $e_i = 0$ and $e_{i+1} \lesssim e^*$, and for the star the theory predicts $e_c = 0$ and $e_p \lesssim e^*$.

To define the admissible downward deviations in the star periphery and the circle more concrete, we numerically simulated the maximum range of deviations from the selfish maximum contribution level, $[e^* - \underline{\eta}\, ,\, e^* + \overline{\eta}]$, which still supports the equilibrium characterizations of Propositions 2-4. For this purpose, we started from payoff function (7) with $e^* = 12$, fixed the values for $\underline{\eta}$ and $\overline{\eta}$ to 1, and checked every possible contribution profile for whether it satisfies the best response criteria according to Definition 1 for all players. We then gradually increased the values for $\underline{\eta}$ and $\overline{\eta}$ until at least one of the statements in Propositions 2-4 failed to be true. Our simulations showed that the binding conditions are those of Proposition 3 on the unique equilibrium in the star. However, as long as the players' individual maximum contribution levels lie in the symmetric interval $[12 - 2, 12 + 2]$, the unique equilibrium is still the periphery-sponsored public good.[15] These margins translate into admissible contribution levels, deviations from best responses, and deviations from neighboring contributions, as shown in Table 2.

(Table 2 about here)

For the distributed equilibria in the dyad, the complete interaction structure, and the circle, the interval $[12-2, 12+2]$ implies that every player's contribution is predicted to lie between $[5\,,\,7]$ in the dyad, $[2.5\,,\,3.5]$ in the complete interaction structure, and $[0.66\,,\,7.33]$ in the circle.[16] Figure 2 (Table 4) shows that the most frequently chosen contributions are 6 for the dyad, 3 for the complete interaction structure, and 4 for the circle. This combination is only predicted by marginal fairness theory. In addition, contributions within the marginal fairness intervals reflect the behavior of the majority of subjects in our experiment: 71.25% of the decisions in the dyad (5-7 units) and 60.63% in the complete interaction structure (2-4 units).[17] As already mentioned in our discussion of selfishness theory, subjects indeed coordinate on a periphery-sponsored star. However, marginal fairness theory is additionally able to explain the peripheral players' downward deviation from $e^* = 12$ (see Fig. 2): 67.22% of their contributions fall into the interval $[10\,,\,12]$. Moreover, our experimental findings for the circle can be much more easily reconciled with marginal fairness concerns as the theory allows for the two types of observed equilibria. In fact, 31.05% of the contributions in Figure 2 can be reconciled with the equilibrium $e_i = 0$ and $e_{i+1} \in [10\,,\,12)$, while another 65.40% are in line with the distributed equilibrium $e_i \in [0.66\,,\,7.33]$.

The evidence produced in Figure 3 is also supportive: 91.67% of all decisions in the dyad, 73.75% in the complete interaction structure, 71.03% in the circle, 73.61% in the star periphery, and 62.5% of all decisions fall into the intervals defined in the $D_i(BR_i)$ column of Table 2.

Finally, marginal fairness theory receives additional support from Figure 4: as predicted for the star, $-12 \leq D_c(e_p) \leq -10$ and $+12 \geq D_p(e_c) \geq +10$ holds in 60.28% of all cases. For the circle, 87.72% of the decisions satisfy either $10 \leq |D_i(e_j)| \leq 12$ or $-6.66 \leq D_i(e_j) \leq +6.66$. Also in the dyad (70.83%) and the complete interaction structure (51.11%) the majority of subjects do not deviate by more than one unit from their neighbors.

Overall marginal fairness theory predicts a large proportion of observed behavior, which is owing to the fact that it combines (a) the predictive power of fairness theories for symmetric and transitive structures (without having to ignore the fact that subjects prefer total contributions of close to 12), with (b) the predictive power of selfishness theory for asymmetric structures, and (c) the coexistence of two possible equilibria in symmetric but intransitive structures.

[15] There are alternative asymmetric, integer-valued intervals sufficient to support the unique equilibrium in the star: $[12-2, 12+1]$ and$[12-1, 12+3]$. In the following, we focus on the symmetric interval.

[16] The interval of individual contributions for the circle is obtained by substituting $\underline{\eta} = \overline{\eta} = 2$ into the first bullet point of Proposition 4.

[17] We rounded the interval boundaries in Table 2 to our advantage.

5.5 Evidence for power- and need-based fairness norms

The theory of power-based norms claims that players in more central positions deserve higher payoffs because a higher reward is intrinsically related to their powerful position. For all but the star, this theory predicts the same outcomes as the theory of pure fairness concerns. For the star the prediction is identical to selfishness theory: a periphery-sponsored local public good (see evidence above). Nevertheless, for the dyad and the complete interaction structure, the theory fails to explain why subjects tend to coordinate on a total contribution close to 12 and systematically deviate downward from $e^* = 12$ in the star periphery position. Moreover, the theory cannot explain why, in the circle where all players are ex ante in the same position, a specialized equilibrium is sometimes played (see our discussion of the selfishness and marginal fairness theories in Sections 5.2 and 5.4).

Finally, the theory of need-based norms, according to which a player in a superior position (higher degree) is expected to help those in inferior positions (lower degree), predicts the same outcomes as the theory of pure fairness concerns. The exception is the star, where it predicts $D_c(e_p) > 0$ and $D_p(e_c) < 0$. The evidence presented above, showing that the star center players (a) never contribute more than 5 units in 95.83% of all their decisions (see Fig. 2 and Table 4) and (b) contribute less than the star periphery players in 92.5% of all their decisions (see Fig. 4 and Table 6), clearly rules out this theory.

Overall, we find that the theory of marginal fairness concerns is closer to our experimental findings than any of the other theories put to the test. Even though each of the other theories (pure selfishness, pure fairness, power- and need-based norms) is able to explain some of our findings, they each fail at least for one of the four interaction structures in our experiment.

6 Robustness

In our theoretical analysis presented in Section 3, we made a number of simplifications: we assumed players to be moderately heterogeneous, we omitted the analysis of three of the interaction structures administered in the experiment, and we limited the analysis to a utility function that represents "difference aversion," thereby neglecting other prominent models of social preferences. We did this to limit the analysis and confine it to the most important insights. In this section we address each of these simplifications in the light of our experimental results.

6.1 Experimental evidence on player heterogeneity

If players were highly heterogeneous with respect to the strength of their fairness concerns or to their fairness norms, then our theoretical predictions, specifically regarding our preferred theory of marginal fairness concerns, would not apply. Instead, an appropriate theory would need to allow for various combinations of more or less extreme fairness concerned players, i.e., combinations of values for α_i and β_i that may be outside the interval defined in Definition 1. Is such an enriched theory necessary to interpret the data?

To assess subjects' heterogeneity, we looked at their deviations from best response play across our entire set of experiments. As best response behavior corresponds to selfish play, difference aversion theory suggests to interpret any deviation as an expression of the importance of other-regarding preferences. Table 3 presents average deviations from best response play (in percentiles) per subject and interaction structure. In all four interaction structures, at least 90% of all subjects invest less than what is required by best response with a minimum of -3.375 and a maximum of 0.708. Thus, as expected from marginal fairness theory, subjects systematically deviate from best response play and typically with a downward deviation. This is in line with observations from previous literature that people feel envy more than they feel guilt (Fehr and Schmidt, 1999). (Table 3 about here)

Furthermore, it is informative to compare the numbers in Table 3 with the intervals for deviation from best response play in Table 2, which support the equilibria of Propositions 2-4. Our calculations for the star periphery show that any deviation from $e^* = 12$ within the interval $[-2\,,\,0]$ can still be classified as evident for marginal fairness concerns. Table 3 shows that this applies to at least 80% of all subjects playing the star. Similarly in the circle, approximately 80% of the subjects did not deviate outside the interval $[-2\,,\,0]$, which is sufficient to support an unequal-contribution equilibrium. Finally, more than 80% and 90% of the subjects in the dyad and the complete interaction structure, respectively, did not systematically deviate outside the interval $[-2\,,\,+2]$. Hence, it is fair to say that about 80% of the subjects in our experiment can be classified as being only marginally fairness concerned.

6.2 Experimental evidence on other network structures

As a second test of our theories, we extend our analysis to the remaining three interaction structures we administered in our experiment. We restrict this test to the two theories with the best predictive power, marginal fairness and the power-based norm. The experimental findings for the core, the dbox, and the line structure are presented in the Appendix.

For the core interaction structure the power-based norm predicts a range of equilibria, in each of which the core center player might make a positive contribution as long as she earns the highest payoff, while the peripheral player invests $e_p = e^* - e_c = 12 - e_c$. The two core players not connected to the peripheral player (core duo) contribute respectively do not contribute as long as they invest the same amount. In the dbox a contribution profile constitutes a power-norm equilibrium if the two connected players make identical contributions, while the players on the edges each contribute $e_p = 12 - 2e_c$ such that the connected players earn more. The same holds for any equilibrium on the line structure.

Marginal fairness concerns will express themselves in equilibria in the core and the d-box, which are subsets of the sets of power-norm equilibria. These subsets can be most clearly identified in the data, which is evident from Figures 5 to 7 in the Appendix: subjects in the core center position contribute nothing, the peripheral subjects contribute approximately 12, and the core duo shares a total contribution of approximately 12. In the d-box the connected players contribute nothing, while the players on the edges make positive contributions, often close to 12. For the line interaction structure, marginal fairness theory predicts potentially unequal contribution levels of the players in the middle, each being strictly smaller than 6, whereas the end players fill the gaps between their neighbors' contributions and approximately 12. Hence, the set of equilibria does not fully coincide with the set of power-norm equilibria. Figures 5 to 7 in the Appendix show that the predictions of both theories are for the most part reflected in our data on the line interaction structure. However, the peak at zero deviation from best response for the subjects in the line middle position as well as the lack of a peak at zero deviation from the neighbors' contributions provides more support for marginal fairness theory.

6.3 Alternative specifications of utility

In our specification of utility function (2), we assume that players compare themselves only with their direct neighbors. Would our predictions for the dyad, the complete interaction structure, the star, and the circle change, if players compared their payoffs with all other players' payoffs in the same game, as, e.g., suggested by Bolton and Ockenfels (2000)? The answer is no. In particular, the predictions for marginally fairness concerned players, summarized in Propositions 2-4, remain unaffected. This is obvious for the complete interaction structure and the dyad as a global comparison is equivalent to a neighborhood comparison. For the star and the circle, the situation is more complicated. However, a global payoff comparison can be easily implemented in the definition of marginally fairness concerned best response play (Definition 1). Player i has an unambiguously larger (smaller) maximum contribution level than the selfish benchmark level only if he or she obtains a higher (lower) payoff than any other player in the game:

$$\begin{aligned}&if\, \pi_i \geq (\leq)\ \pi_j \text{ for all } j \in N\backslash\{i\},\\ &\qquad \text{with at least one inequality being strict, then } e_i^* > (<)\ e^*.\end{aligned}$$

Working with this adjusted definition has no implications for the proof of Proposition 4 on the circle. The proof of Proposition 3 on the star needs to be adjusted in two respects: in the last line of point one, it does not necessarily follow that $e_j^* > e^*$ because player j might attain a higher payoff than another peripheral player k. However, there is necessarily a peripheral player l who earns weakly more than anybody else, including the star center. This implies that $e_l^* > e^*$, which leads to a contradiction to $e_i^* < e^*$. A similar argument can be made for the last line of point two.

Next to the "difference aversion models" studied in Section 3, so-called "social welfare models" assume that "people like to increase social surplus, caring especially about helping those (themselves or others) with low payoffs" (Rabin and Charness, 2002, p.818). These models incorporate concerns for efficiency, where an individual player's utility is a convex combination of own well-being and social welfare (Charness and Rabin, 2002):

$$U_i = \gamma\pi_i + (1-\gamma)\left[\delta min_{k\in N_i(g)\cup i}\pi_k + (1-\delta)\sum_{j\in N_i(g)}\pi_j\right].$$

It is obvious that strong efficiency concerns would raise a player's personal maximum contribution level, as she internalizes the social gains of an additional contribution in the neighborhood. Hence, we would expect to see a tendency toward a positive deviation from selfish best response play in the data, which is clearly refuted in Figure 3 (and Fig. 6 for the other three structures).

7 Conclusions

The interpretation of evidence from previous experiments has been that "a substantial percentage of the people are strongly motivated by other-regarding preferences" (Fehr and Schmidt, 2006, p. 615), leading to the conclusion that concerns for the well-being of others, triggered by fairness or reciprocity, cannot be ignored in social and economic interactions. Fehr and Schmidt (2006) conclude that "the real question is no longer whether many people have other-regarding preferences, but under which conditions these preferences have important economic and social effects." Our aim has been to contribute to answering this question and, in particular, to develop a parsimonious utility theory of marginal fairness able to explain decision making in various laboratory settings. The paper also discusses findings from an experimental evaluation of this theory.

By imposing an interaction structure to a local public good decision situation and systematically varying this structure along two dimensions, symmetry and transitivity, we have been able to model various degrees of influence between individuals. In line with previous experimental studies, we find that outcomes are rather equitable in the extreme case of homogeneous influence between all individuals. A typical example for this decision situation is the complete interaction structure as, e.g., in the classic public goods game with homogeneous players (as, e.g., studied in Isaac and Walker, 1988a). For the intermediate case with symmetric influence across individuals - but where an individual's actions do not influence all others (intransitivity) - we find equitable as well as rather inequitable outcomes. Such structures refer, e.g., to symmetric buyer-seller negotiations among homogeneous individuals. At the other extreme, where some individuals' actions affect many others, while other individuals only affect a few, we find that rather inequitable outcomes are observed with higher frequency. This is in line with previous experimental evidence using asymmetric and intransitive settings such as, e.g., in multilateral bargaining and other star-like structures (e.g. Schotter, Weiss, and Zapater (1996), Roth, Prasnikar, Fujiwara, and Zamir (1991),Fehr and Fischbacher, 2002).

We tested several established fairness theories (stable inequity aversion or social welfare concerns, power- and need-based norms) as well as pure selfishness as a benchmark case against the marginal fairness theory introduced in this paper, in which other-regarding concerns are assumed to be present but weak. Overall, our experimental observations are best explained by our preferred theory of marginal fairness concerns.

In our experiment, we did not consider settings where individuals' influence is heterogeneous but each individual influences all others (as, e.g., in Cournot games or heterogeneous public goods games). Previous studies indicate that in these settings outcomes are also less than equitable (see e.g.Buckley and Croson, 2006, Maurice, Rouaix, and Willinger, 2013). Based on these observations and on our own theoretical analysis, we have come to several conclusions. First, by taking into account the (a)symmetry and the (in)transitivity of the interaction structure between individuals, previous experimental results on the strength of fairness concerns can be better understood. Second, fairness concerns are, on average, only marginal in the sense that behavior is rather similar to selfish best response play. Only in specific (fully symmetric and transitive) situations do fairness concerns help to coordinate on equitable outcomes. Finally, for highly asymmetric and intransitive structures, marginal fairness concerns do not preclude the existence of unfair outcomes and may even favor extremely unfair behavior.

Our study and findings relate to the existing literature on what individuals regard as fair in a given situation. Kahneman, Knetsch, and Thaler (1991) introduced different distributive norms into the discussion, and in the past 20 years a substantial amount of research has been devoted to the understanding the nature of social preferences.[18] Fehr and Fischbacher (2002) present evidence on the effect of competition on the frequency of fair outcomes in the Ultimatum Game. While we agree with the authors' point that economists fail to understand core questions in economics if they insist on the self-interest hypothesis, our findings qualify their statement that the interaction between material incentives and social preferences is likely to have important effects. Our results suggest that the vast majority of individuals in traditional experimental settings only care about others' well-being when the cost of deviating from their selfishly optimal decision is comparatively small, independent of the decision situations they are confronted with.

[18] See the surveys by Rabin (1993), Rabin and Charness (2002), Levitt and List (2007), and Fehr and Schmidt (1999).

Author affiliation:

Tahera Rezaei Khavas, Stephanie Rosenkranz, and Bastian Westbrock
Utrecht University School of Economics
Kriekenpitplein 21-22
3584 EC Utrecht
The Netherlands

Utz Weitzel
Radboud University Nijmegen
Department of Economics
Thomas van Aquinostraat 5
6525 GD Nijmegen
The Netherlands

Appendix

Instructions

(Experimental instructions about here)

Additional tables

(Table 4 about here)

(Table 5 about here)

(Table 6 about here)

Treatments 5-7 (line, core, dbox)

(Figure 5 about here)

(Figure 6 about here)

(Figure 7 about here)

References

ANDREONI, J., W. HARBAUGH, AND L. VESTERLUND (2003): "The carrot or the stick: Rewards, punishments, and cooperation," *The American Economic Review*, 93(3), 893–902. 4

ANDREONI, J., AND J. MILLER (2002): "Giving according to GARP: An experimental test of the consistency of preferences for altruism," *Econometrica*, 70(2), 737–753. 3, 4

BINMORE, K. (2005): *Natural justice*. Oxford University Press. 1, 3

BLANCO, M., D. ENGELMANN, AND H. T. NORMANN (2011): "A within-subject analysis of other-regarding preferences," *Games and Economic Behavior*, 72(2), 321–338. 4

BOLTON, G. E., AND A. OCKENFELS (2000): "ERC: A theory of equity, reciprocity, and competition," *American Economic Review*, pp. 166–193. 3, 6.3

BRAMOULLÉ, Y., AND R. KRANTON (2007): "Public goods in networks," *Journal of Economic Theory*, 135(1), 478–494. 1, 2, 2, 3.1, 3.1, 3.3.1, 3.4, 3.5

BUCKLEY, E., AND R. CROSON (2006): "Income and wealth heterogeneity in the voluntary provision of linear public goods," *Journal of Public Economics*, 90(4), 935–955. 7

CHARNESS, G., AND M. RABIN (2002): "Understanding social preferences with simple tests," *The Quarterly Journal of Economics*, 117(3), 817–869. 6.3

COOK, K. S., AND R. M. EMERSON (1978): "Power, Equity and Commitment in Exchange Networks," *American Sociological Review*, 43(5), 721. 1

DAVIS, D. D., AND C. A. HOLT (1993): *Experimental Economics, 1993*. Princeton Univ. Press, Princeton. 1

FALK, A., AND M. KOSFELD (2012): "It's all about connections: Evidence on network formation," *Review of Network Economics*, 11(3). 4

FEHR, E., AND A. FALK (1999): "Wage rigidity in a competitive incomplete contract market," *Journal of Political Economy*, 107(1), 106–134. 4

FEHR, E., AND U. FISCHBACHER (2002): "Why social preferences matter–the impact of non–selfish motives on competition, cooperation and incentives," *The Economic Journal*, 112(478), C1–C33. 1, 7

FEHR, E., AND K. M. SCHMIDT (1999): "A Theory of Fairness, Competition, and Cooperation," *The Quarterly Journal of Economics*, 114(3), 817–868. 1, 3, 6.1, 18

——— (2006): "The economics of fairness, reciprocity and altruism − experimental evidence and new theories," *Handbook of the economics of giving, altruism and reciprocity*, 1, 615–691. 1, 7

FISCHBACHER, U. (2007): "z–Tree: Zurich toolbox for ready–made economic experiments," *Experimental Economics*, 10(2), 171–178. 4

FISMAN, R., S. KARIV, AND D. MARKOVITS (2007): "Individual preferences for giving," *The American Economic Review*, pp. 1858–1876. 4

FORSYTHE, R., J. L. HOROWITZ, N. E. SAVIN, AND M. SEFTON (1994): "Fairness in simple bargaining experiments," *Games and Economic Behavior*, 6(3), 347–369. 1

FUDENBERG, D., AND J. TIROLE (1991): *Game Theory*. MIT Press. 3.5

GOEREE, J. K., A. RIEDL, AND A. ULE (2009): "In search of stars: Network formation among heterogeneous agents," *Games and Economic Behavior*, 67(2), 445–466. 4

GREINER, B. (2004): "An online recruitment system for economic experiments," . 4

GÜTH, W., R. SCHMITTBERGER, AND B. SCHWARZE (1982): "An experimental analysis of ultimatum bargaining," *Journal of Economic Behavior & Organization*, 3(4), 367–388. 1

HARRISON, G. W., AND J. HIRSHLEIFER (1989): "An experimental evaluation of weakest link/best shot models of public goods," *The Journal of Political Economy*, pp. 201–225. 1, 3.4

ISAAC, R. M., AND J. M. WALKER (1988a): "Communication and free-riding behavior: The voluntary contribution mechanism," *Economic Inquiry*, 26(4), 585–608. 7

——— (1988b): "Group size effects in public goods provision: The voluntary contributions mechanism," *The Quarterly Journal of Economics*, 103(1), 179–199. 1

ISAAC, R. M., AND J. M. WALKER (1998): "Nash as an organizing principle in the voluntary provision of public goods: Experimental evidence," *Experimental Economics*, 1(3), 191–206. 6

JOUNG, Y.-J., T. H.-Y. CHIU, AND S.-M. CHEN (2012): "Cooperating with free riders in unstructured P2P networks," *Computer Networks*, 56(1), 198–212. 1

KAHNEMAN, D., J. L. KNETSCH, AND R. H. THALER (1991): "Anomalies: The endowment effect, loss aversion, and status quo bias," *The Journal of Economic Perspectives*, 5(1), 193–206. 7

LAURY, S. K., AND C. A. HOLT (1998): "Voluntary provision of public goods: Experimental results with interior Nash equilibria," *Handbook of Experimental Economic Results*. 6

LAURY, S. K., J. M. WALKER, AND A. W. WILLIAMS (1999): "The voluntary provision of a pure public good with diminishing marginal returns," *Public Choice*, 99(1-2), 139–160. 6

LAWLER, E. J., AND J. YOON (1998): "Network Structure and Emotion in Exchange Relations," *American Sociological Review*, 63(6), 871–894. 1

LEVITT, S. D., AND J. A. LIST (2007): "What do laboratory experiments measuring social preferences reveal about the real world?," *The Journal of Economic Perspectives*, pp. 153–174. 18

LIST, J. A. (2007): "On the interpretation of giving in dictator games," *Journal of Political Economy*, 115(3), 482–493. 1

LOEWENSTEIN, G. F., L. THOMPSON, AND M. H. BAZERMAN (1989): "Social utility and decision making in interpersonal contexts.," *Journal of Personality and Social Psychology*, 57(3), 426. 3

LOVAGLIA, M. J., J. SKVORETZ, D. WILLER, AND B. MARKOVSKY (1995): "Negotiated Exchanges in Social Networks," *Social Forces*, 74(1), 123. 1

MARWELL, G., AND R. E. AMES (1980): "Experiments on the provision of public goods. II. Provision points, stakes, experience, and the free–rider problem," *The American Journal of Sociology*, 85(4), 926–937. 1

——— (1981): "Economists free ride, does anyone else?: Experiments on the provision of public goods, IV," *Journal of Public Economics*, 15(3), 295–310. 1

MAURICE, J., A. ROUAIX, AND M. WILLINGER (2013): "Income Redistribution and Public Good Provision: An Experiment," *International Economic Review*, 54(3), 957–975. 7

MOLM, L. D., G. PETERSON, AND N. TAKAHASHI (1999): "Power in Negotiated and Reciprocal Exchange," *American Sociological Review*, 64(6), 876. 1

PLOTT, C. R. (1982): "Industrial organization theory and experimental economics," *Journal of Economic Literature*, 20(4), 1485–1527. 1, 3.4

——— (1983): "Externalities and corrective policies in experimental markets," *The Economic Journal*, pp. 106–127. 1

RABIN, M. (1993): "Incorporating fairness into game theory and economics," *The American Economic Review*, pp. 1281–1302. 7, 18

RABIN, M., AND G. CHARNESS (2002): "Understanding Social Preferences with Simple Tests," *The Quarterly Journal of Economics*, pp. 817–869. 1, 6.3, 18

RAMZAN, N., H. PARK, AND E. IZQUIERDO (2012): "Video streaming over P2P networks: Challenges and opportunities," *Signal Processing: Image Communication*, 27(5), 401–411. 1

ROSENKRANZ, S., AND U. WEITZEL (2012): "Network structure and strategic investments: An experimental analysis," *Games and Economic Behavior*, 75(2), 898–920. 4

ROTH, A. E., V. PRASNIKAR, M. O. FUJIWARA, AND S. ZAMIR (1991): "Bargaining and Market Behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An Experimental Study," *The American Economic Review*, 81(5), 1068–1095. 1, 3.4, 7

SCHOTTER, A., A. WEISS, AND I. ZAPATER (1996): "Fairness and survival in ultimatum and dictatorship games," *Journal of Economic Behavior & Organization*, 31(1), 37–56. 7

SEFTON, M., AND R. STEINBERG (1996): "Reward structures in public good experiments," *Journal of Public Economics*, 61(2), 263–287. 6

SMITH, V. L. (1962): "An experimental study of competitive market behavior," *The Journal of Political Economy*, pp. 111–137. 1

——— (1979): "An experimental comparison of three public good decision mechanisms," *Scandinavian Journal of Economics*, 81(2), 198–215. 1

——— (1980): "Experiments with a decentralized mechanism for public good decisions," *The American Economic Review*, pp. 584–599. 1

SOBEL, J. (2005): "Interdependent Preferences and Reciprocity," *Journal of Economic Literature*, 43(2), 392–436. 1

WILLINGER, M., AND A. ZIEGELMEYER (2001): "Strength of the social dilemma in a public goods experiment: An exploration of the error hypothesis," *Experimental Economics*, 4(2), 131–144. 6

Structure/ Theory	Pure fairness	Pure selfishness	Marginal fairness	Power norm	Need norm
Dyad	$e_1 = e_2 \geq 0$	$e_1, e_2 \geq 0$ s.t. $e_1 + e_2 = e^*$	$e_1 = e_2 \approx \frac{e^*}{2}$	see pure fairness	see pure fairness
Complete	$e_i = e_j \geq 0$ $\forall i, j \in N$	$e_i \geq 0$ s.t. $\sum_{i \in N} e_i = e^*$	$e_i = e_j \approx \frac{e^*}{N}$ $\forall i, j \in N$	see pure fairness	see pure fairness
Star	$e_c \geq e_p$ s.t. $\pi_c = \pi_p$	1) $e_c = 0$, (!) $e_p = e^*$ 2) $e_c = e^*$, $e_p = 0$	$e_c = 0$, $e_p \lessapprox e^*$	$e_c = 0$, $e_p = e^*$	$e_c = e^*$, $e_p = 0$
Circle	1) $e_i = e_j \geq 0$ $\forall i, j \in N$ 2) $e_i > e_j = e_l > e_k$ s.t. $\pi_i = \pi_j \; \forall i, j \in N$	1) $e_i = e_j = \frac{e^*}{3}$ 2) $e_i = 0$, $e_{i+1} = e^*$ (!)	1) $e_i \approx \frac{e^*}{3}$ 2) $e_i = 0$, $e_{i+1} \lessapprox e^*$	see pure fairness	see pure fairness

Table 1: Equilibria for different theories and interaction structures

Node type	e_i		$D_i(BR_i)$		$D_i(e_j)$	
	Min	Max	Min	Max	Min	Max
Dyad	5	7	-2	+2	0	0
Complete	2.5	3.5	-2	+2	0	0
Star periphery	10	12	-2	0	10	12
Star center	0	0	0	0	-12	-10
Circle, specialized i	10	12	-2	0	10	12
Circle, specialized i+1	0	0	0	0	-12	-10
Circle, distributed	0.66	7.33	-2	+2	-6.66	+6.66

Table 2: Intervals for individual contributions, e_i, deviation from best response, $D_i(BR_i)$, and deviation from contribution of neighbors, $D_i(e_j)$, supporting equilibria of Propositions 2-4

Percentiles	Dyad	Star	Circle	Complete
1%	-3.250	-3.125	-3.375	-3.250
5%	-2.729	-2.583	-2.625	-2.500
10%	-2.292	-2.000	-2.021	-1.917
25%	-1.458	-1.500	-1.396	-1.417
50%	-1.063	-1.083	-0.896	-0.917
75%	-0.625	-0.688	-0.625	-0.458
90%	-0.188	-0.063	-0.146	-0.083
95%	0.083	0.271	0.313	0.313
99%	0.417	0.542	0.708	0.708
Obs	120	120	120	120
Mean	-1.142	-1.113	-1.023	-0.999
Std. Dev.	0.804	0.788	0.833	0.807

Table 3: Average deviations from best response per subject

Table 4: Individual contribution (in percent)

Contribution	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	29	Total
Node type																		
Dyad	2.50	3.13	2.08	3.75	7.08	14.58	42.50	14.17	5.00	0.83	0.63	1.25	1.67	0.42	0.21	0.00	0.21	100
Complete	15.42	11.88	22.50	26.67	11.46	5.21	1.67	1.04	1.04	1.25	0.21	0.42	0.83	0.21	0.00	0.21	0.00	100
Star: center	62.50	15.00	7.50	5.00	3.33	2.50	0.83	2.50	0.00	0.83	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100
Star: periphery	1.39	0.83	1.39	2.22	0.83	4.44	3.33	3.33	3.89	11.11	12.78	13.61	39.72	1.11	0.00	0.00	0.00	100
Circle	19.17	8.75	8.75	12.08	19.38	6.46	5.42	3.33	1.88	1.67	2.71	2.29	6.88	0.83	0.21	0.21	0.00	100
Line: end	2.92	1.67	6.25	6.25	7.50	5.42	13.33	7.50	7.92	6.25	6.67	11.67	15.83	0.00	0.42	0.42	0.00	100
Line: middle	25.42	11.67	15.83	13.33	11.67	7.50	5.83	2.92	3.33	0.83	0.00	0.42	0.42	0.83	0.00	0.00	0.00	100
Core: center	68.33	11.67	13.33	5.00	0.00	1.67	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100
Core: periphery	1.67	2.50	1.67	1.67	5.00	3.33	4.17	3.33	4.17	5.83	9.17	14.17	40.83	2.50	0.00	0.00	0.00	100
Core: duo	8.75	8.75	6.67	9.58	11.67	13.33	17.50	7.50	5.83	2.50	2.92	2.92	2.08	0.00	0.00	0.00	0.00	100
Dbox: connected	44.58	16.25	10.83	15.00	5.42	2.92	2.50	0.42	0.83	0.42	0.42	0.00	0.42	0.00	0.00	0.00	0.00	100
Dbox: edges	5.00	5.00	7.92	10.42	8.33	9.58	7.50	5.42	5.42	5.42	6.67	7.08	14.58	1.25	0.42	0.00	0.00	100

Table 5: Deviation $D_i(BR_i)$ from best response (in percent)

$D_i(BR_i)$	-12	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	9	12	24	Total
Node type																								
Dyad	0.00	0.42	0.42	0.42	0.00	1.25	1.67	1.25	3.33	7.92	10.00	10.00	41.25	6.88	10.83	2.71	0.42	0.42	0.00	0.21	0.00	0.42	0.21	100
Complete	0.00	0.00	1.67	0.00	0.83	1.67	3.33	2.50	8.33	11.67	12.50	14.17	23.96	9.58	5.21	3.13	1.04	0.00	0.00	0.42	0.00	0.00	0.00	100
Star: center	0.00	0.00	0.00	0.00	0.00	0.83	0.00	0.00	0.00	0.00	0.00	0.83	62.50	15.83	6.67	5.83	2.50	2.50	0.00	2.50	0.00	0.00	0.00	100
Star: periphery	0.56	0.56	0.28	0.83	0.83	1.11	2.50	1.94	2.78	9.17	10.28	7.22	56.11	4.17	1.11	0.28	0.28	0.00	0.00	0.00	0.00	0.00	0.00	100
Circle	0.00	0.00	0.00	0.63	2.08	1.25	2.50	3.96	7.71	7.92	11.88	10.63	35.63	7.50	4.79	2.08	0.42	0.83	0.21	0.00	0.00	0.00	0.00	100
Line: end	0.00	0.83	1.25	0.42	0.83	3.33	2.08	5.00	7.08	9.58	15.00	14.17	34.17	3.75	1.67	0.42	0.42	0.00	0.00	0.00	0.00	0.00	0.00	100
Line: middel	0.00	0.00	0.42	0.42	0.42	2.50	0.42	2.50	2.08	2.50	6.67	10.83	39.17	17.50	7.92	4.58	1.67	0.00	0.00	0.00	0.42	0.00	0.00	100
Core: center	0.00	0.83	0.00	0.00	0.83	0.00	2.50	0.00	0.00	0.00	0.83	1.67	67.50	12.50	11.67	0.83	0.00	0.83	0.00	0.00	0.00	0.00	0.00	100
Core: periphery	0.83	0.00	0.00	3.33	4.17	3.33	2.50	2.50	5.83	5.00	8.33	11.67	49.17	2.50	0.83	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100
Core: duo	0.83	0.83	0.83	0.83	0.83	1.67	1.67	5.00	10.00	6.67	13.33	15.83	36.67	3.33	1.67	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100
Dbox: connected	0.00	0.00	0.00	0.00	0.00	1.67	0.00	1.67	1.67	1.67	2.50	11.67	45.42	16.25	7.50	6.67	1.25	1.67	0.00	0.42	0.00	0.00	0.00	100
Dbox: edge	0.42	0.42	0.00	1.25	3.33	5.42	3.33	6.67	10.83	8.75	15.83	8.75	29.17	3.75	1.25	0.83	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100

Table 6: Deviation $D_i(e_j)$ from contribution of neighbors by type of node position (in percent)

$D_i(e_j)$	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	22
Node type																	
Dyad	49.17	10.83	3.75	2.92	1.88	0.42	0.63	0.83	0.63	0.42	1.25	0.21	1.04	0.21	0.21	0.00	0.21
Complete	19.86	15.63	10.00	5.97	2.50	1.25	1.04	0.69	0.56	0.76	0.21	0.49	0.63	0.14	0.07	0.14	0.00
Star: center	2.22	1.39	1.94	0.00	0.83	0.28	0.00	0.28	0.28	0.28	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Star: periphery	2.22	1.67	1.94	1.39	1.67	2.22	6.11	3.06	5.56	7.78	16.67	6.11	37.50	0.83	0.00	0.00	0.00
Circle	14.58	8.23	6.15	4.90	2.29	1.98	2.19	1.88	1.25	1.77	1.98	2.08	6.77	0.83	0.21	0.21	0.00
Line: end	5.83	7.08	5.00	4.58	7.50	5.83	6.67	2.08	5.00	2.08	7.08	7.08	14.17	0.00	0.42	0.42	0.00
Line: middle	7.50	7.71	6.88	2.92	5.63	2.92	1.88	1.04	1.46	0.63	0.21	0.21	0.42	0.42	0.00	0.00	0.00
Core: center	5.83	3.89	3.33	0.56	0.00	0.28	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Core: periphery	0.83	0.00	2.50	3.33	5.83	2.50	3.33	3.33	6.67	4.17	9.17	9.17	41.67	1.67	0.00	0.00	0.00
Core: duo	13.33	7.50	7.71	4.38	9.58	7.08	9.38	4.17	3.75	1.88	2.29	2.08	1.88	0.00	0.00	0.00	0.00
Dbox: connected	16.67	7.78	4.72	2.78	2.36	1.25	0.56	0.42	0.56	0.42	0.28	0.00	0.14	0.00	0.00	0.00	0.00
Dbox: edges	6.67	6.04	6.46	7.08	6.04	6.04	5.83	3.13	4.79	4.58	6.67	5.00	14.17	1.25	0.21	0.00	0.00

$D_i(e_j)$	-22	-15	-14	-13	-12	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0
Node type																	
Dyad	0.21	0.00	0.21	0.21	1.04	0.21	1.25	0.42	0.63	0.83	0.63	0.42	1.88	2.92	3.75	10.83	49.17
Complete	0.00	0.14	0.07	0.14	0.63	0.49	0.21	0.76	0.56	0.69	1.04	1.25	2.50	5.97	10.00	15.63	19.86
Star: center	0.00	0.00	0.00	0.83	37.50	6.11	16.67	7.78	5.56	3.06	6.11	2.22	1.67	1.39	1.94	1.67	2.22
Star: periphery	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.28	0.28	0.28	0.00	0.28	0.83	0.00	1.94	1.39	2.22
Circle	0.00	0.21	0.21	0.83	6.77	2.08	1.98	1.77	1.25	1.88	2.19	1.98	2.29	4.90	6.15	8.23	14.58
Line: end	0.00	0.00	0.00	0.42	0.42	0.42	0.00	0.42	0.42	0.83	1.67	0.83	4.17	0.42	3.33	5.83	5.83
Line: middle	0.00	0.21	0.21	0.21	7.29	3.54	3.75	1.46	3.75	1.67	4.38	5.42	7.29	5.00	7.71	8.33	7.50
Core: center	0.00	0.00	0.00	0.56	15.28	4.72	4.17	3.06	6.39	5.83	10.83	8.89	10.00	5.83	6.11	4.44	5.83
Core: periphery	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.83	1.67	3.33	0.83
Core: duo	0.00	0.00	0.00	0.00	0.83	0.83	1.46	0.63	0.63	0.63	2.08	1.25	3.54	1.04	5.83	6.25	13.33
Dbox: connected	0.00	0.00	0.14	0.83	9.44	3.33	4.58	3.19	3.47	2.22	4.17	4.58	4.86	6.67	6.39	8.19	16.67
Dbox: edges	0.00	0.00	0.00	0.00	0.21	0.00	0.21	0.42	0.42	0.42	0.42	1.04	2.29	1.25	3.96	5.42	6.67

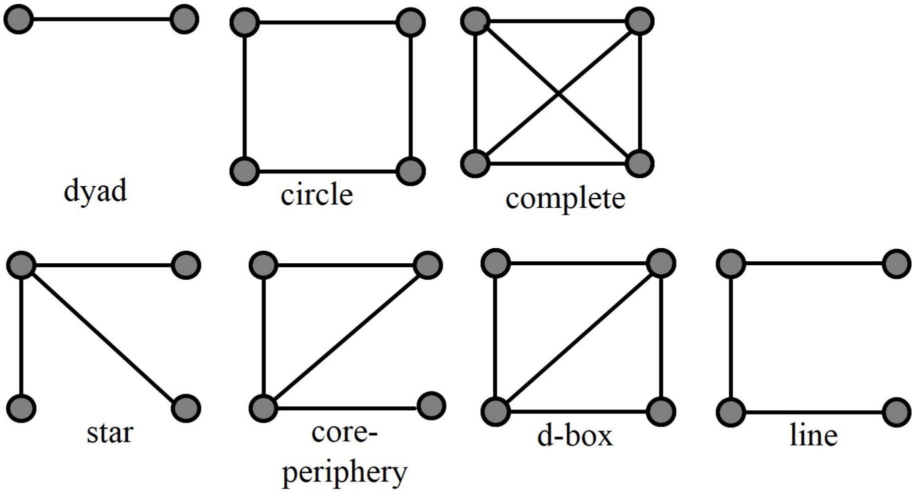


Figure 1: Experimental interaction structures

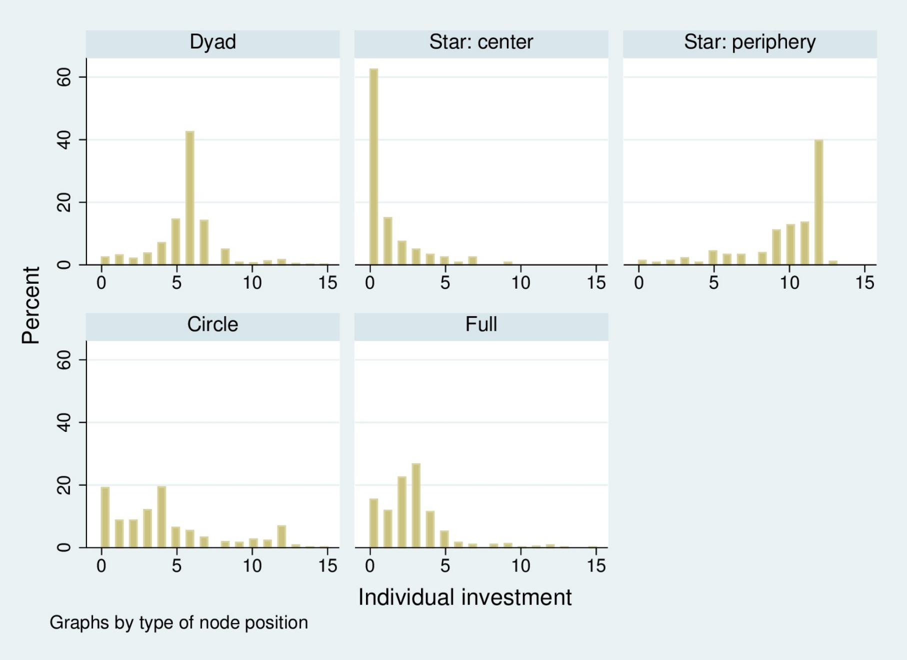


Figure 2: Individual contributions, e_i

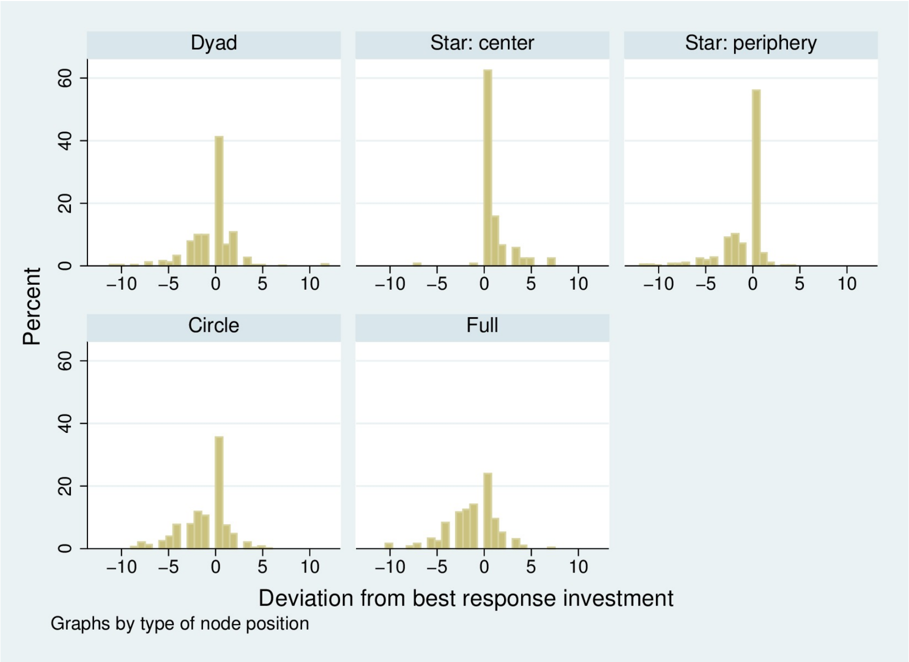


Figure 3: Deviation $D_i(BR_i)$ from best response

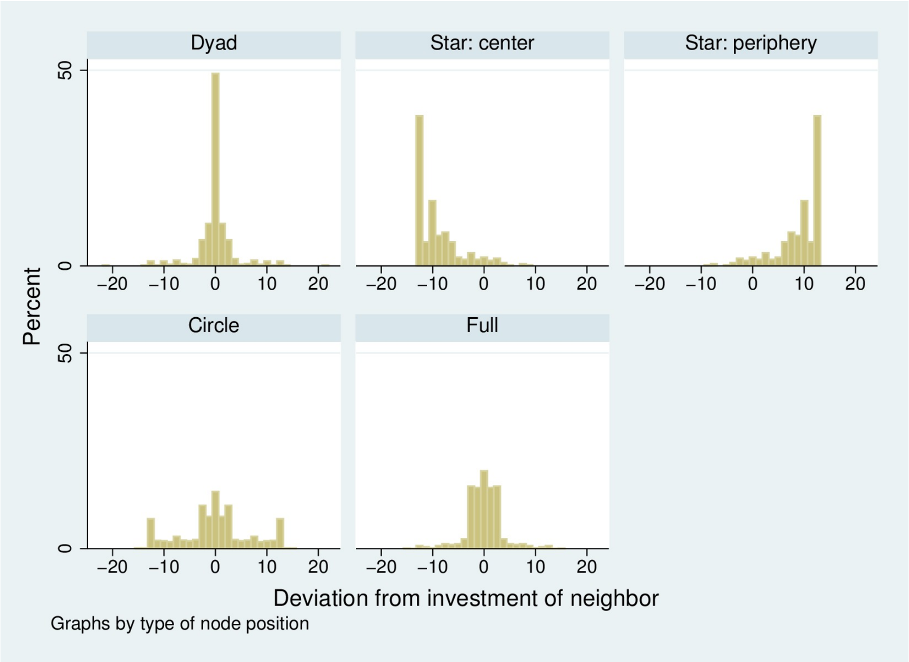


Figure 4: Deviation $D_i(e_j)$ from contribution of neighbors by type of node position

- Instructions -

Please read the following instructions carefully. These instructions state everything you need to know in order to participate in the experiment. If you have any questions, please raise your hand. One of the experimenters will approach you in order to answer your question. The rules are equal for all the participants.

You can earn money by means of earning points during the experiment. The number of points that you earn depends on your own choices, and the choices of other participants. At the end of the experiment, the total number of points that you earn during the experiment will be exchanged at an exchange rate of:

400 points = 1 Euro

The money you earn will be paid out in cash at the end of the experiment without other participants being able to see how much you earned. Further instructions on this will follow in due time. During the experiment you are not allowed to communicate with other participants. Turn off your mobile phone and put it in your bag. Also, you may only use the functions on the screen that are necessary for the functioning of the experiment. Thank you very much.

- Overview of the experiment -

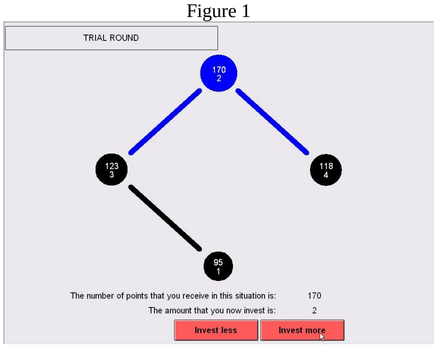
The experiment consists of *seven scenarios*. Each scenario consists again of *one trial round* and *four paid rounds* (altogether 35 rounds of which 28 are relevant for your earnings).

In *all scenarios* you will be *grouped* with either one or with three other randomly selected participants. At the beginning of *each of the 35 rounds*, the groups and the positions within the groups will be randomly changed. The participants that you are grouped with in one round are very likely different participants from those you will be grouped with in the next round. It will not be revealed with whom you were grouped at any moment during or after the experiment.

The participants in your group (of two or four players, depending on the scenario) will be shown as circles on the screen (see Figure 1). You are displayed as a **blue** circle, while the other participants are displayed as **black** circles. You are always connected to one or more other participants in your group. These other participants will be called *your neighbors*. These connections differ per scenario and are displayed as lines between the circles on the screen (see also Figure 1).

Each round lasts *between 30 and 90 seconds*. The end will be at an unknown and random moment in this time interval. During this time interval you can earn **points** by producing know-how, but producing know-how also costs points. The points you receive in the end depend on your own investment in know-how and the investments of your neighbors.

Figure 1



The number of points that you receive in this situation is: 170

The amount that you now invest is: 2

By clicking on one of the two buttons at the bottom of the screen you increase or decrease your investment in know-how. At the end of the round, you receive the amount of points that is shown on the screen at that moment in time. In other words, your final earnings only depend on the situation at the end of every round. Note that this end can be at any between 30 and 90 seconds after the round is started and that this moment is unknown to everybody. Also different rounds will not last equally long.

The points you will *receive* can be seen as the *top number* in your blue circle. The points others will *receive* are indicated as the *top number* in the black circles of others. Next to this, the *size of the circles* changes with the points that you and the other participants will receive: a larger circle means that the particular participant receives more points. The *bottom number* in the circles indicates the amount *invested* in know-how by the participants in your group.

Remarks:

- It can occur that there is a time-lag between your click and the changes of the numbers on the screen. One click is enough to change your investment by one. A subsequent click will not be effective until the first click is effectuated.
- **Therefore wait until your investment in know-how is adapted before making further changes!**

- Your earnings -

Now we explain how the number of points that you earn depends on the investments. Read this carefully. Do not worry if you find it difficult to grasp immediately. We also present an example with calculations below. Next to this, there is a trial round for each scenario to gain experience with how your investment affects your points.

In all scenarios, the points you receive at the end of each round depend in the same way on two factors:

1. **Every unit that you invest in know-how yourself will cost you 5 points**.
2. **You earn points for each unit that you invest yourself and for each unit that your neighbors invest.**

If you sum up all units of investment of yourself and your neighbors, the following table gives you the points that you earn from these investments:

Your investment plus your neighbors' investments	0	1	2	3	4	5	6	7	8	9	10
Points	0	28	54	78	100	120	138	154	168	180	190

Your investment plus your neighbors' investments	11	12	13	14	15	16	17	18	19	20	21
Points	198	204	208	210	211	212	213	214	215	216	217

The higher the total investments, the lower are the points earned from an additional unit of investment. Beyond an investment of 21, you earn one extra point for every additional unit invested by you or one of your neighbors.

Note: if your and your neighbors' investments add up to 12 or more, earnings increase by less than 5 points for each additional unit of investment.

- Example -

Suppose

1. you invest 2 units;
2. one of your neighbors invests 3 units and another neighbor invests 4 units.

Then you have to pay 2 times 5 = 10 points for your own investment.

The investments that you profit from are your own plus your neighbors' investments: 2 + 3 + 4 = 9. In the table you can see that your earnings from this are 180 points.

In total, this implies that you receive 180 − 10 = 170 points if this would be the situation at the end of the round. Figure 1 shows this example as it would appear on the screen. The investment of the fourth participant in your group does not affect your earnings. In the trial round before each of the seven scenarios, you will have time to get used to how the points you will receive change with investments.

- Scenarios -

All rounds are basically the same. The only thing that changes between scenarios is whether you are in a group of two or four participants and how participants are connected to each other. Also your own position randomly changes within scenarios and between rounds. We will notify you each time on the screen when a new scenario and trial round starts. At the top of the screen you can also see when you are in a trial round (see top left in Figure 1). Paying rounds are just indicated by "ROUND" while trial rounds are indicated by "TRIAL ROUND".

- Questionnaire -

After the 35 rounds you will be asked to fill in a questionnaire. Please take your time to fill in this questionnaire accurately. In the mean time your earnings will be counted. Please remain seated until the payment has taken place.

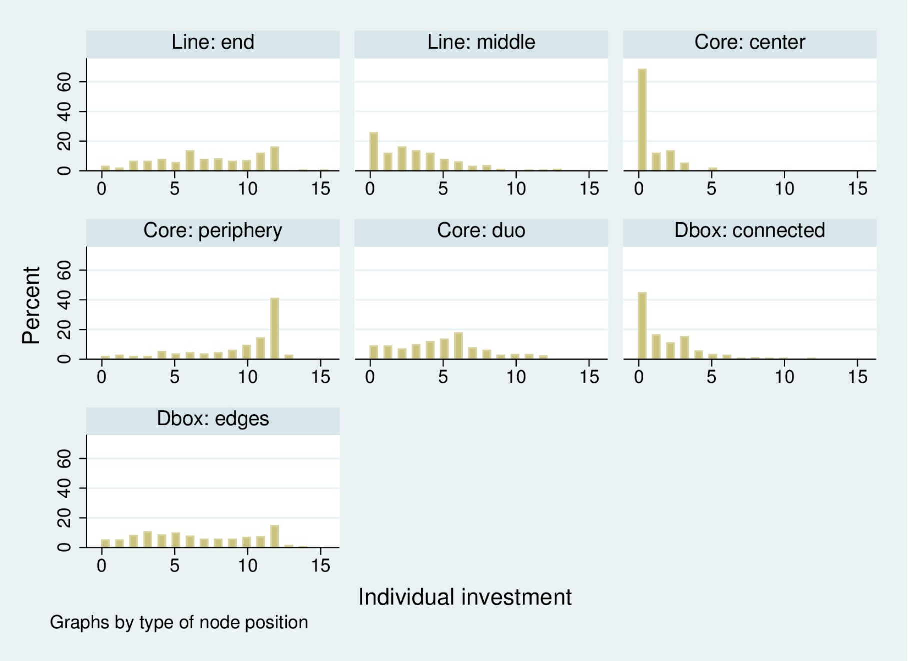


Figure 5: Individual contributions in line, core and dbox

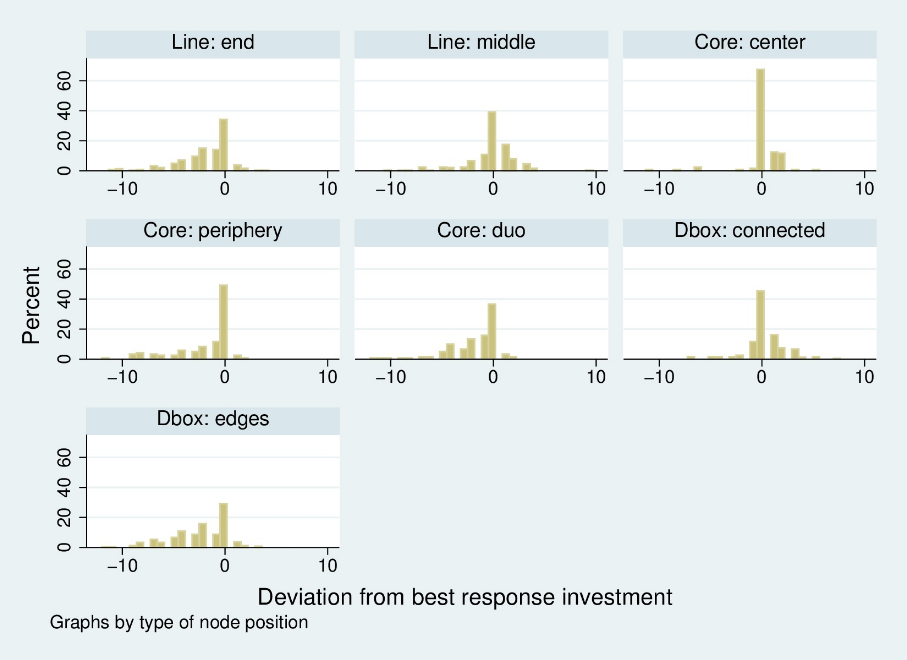


Figure 6: Individual deviation from best response in line, core and dbox

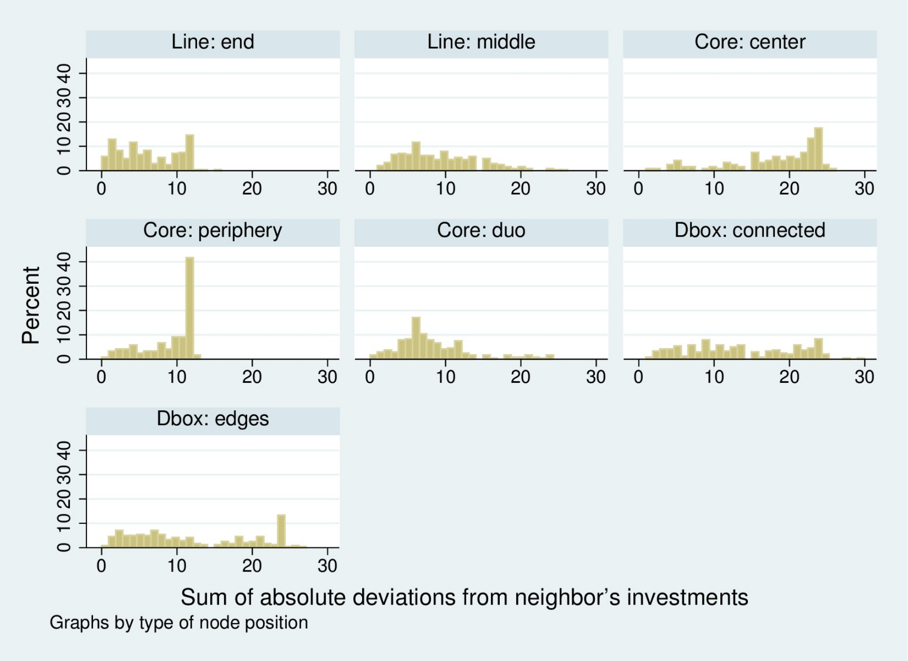


Figure 7: Deviation $D_i(IN_j)$ from contribution of neighbors by type of node position in line, core and dbox