Westbrock, Bastian; Rosenkranz, Stephanie; Rezaei, Sarah; Weitzel, Utz

Conference Paper

Social preferences on networks

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2021: Climate Economics

Provided in Cooperation with:
Verein für Socialpolitik / German Economic Association

Social Preferences on Networks

Sarah Rezaei[*] Stephanie Rosenkranz Utz Weitzel[†]
Bastian Westbrock[‡]

December 15, 2020

Abstract

We develop a model of social preferences for network games and study its predictions in a local public goods game with multiple equilibria. The key feature is that players' social preferences are heterogeneous. This gives room for disagreement between players about the "right" payoff ordering. When preferences are compatible, however, players coordinate on a refined equilibrium set. How easily the requirements for preference compatibility are met crucially depends on a property of the network structure: neighborhood nestedness. This means that equilibrium selection succeeds in small, connected structures but also in centralized networks. All predictions are confirmed in an experiment.

JEL: D85, C70, C91, H41
Keywords: social preferences, network games, equilibrium selection

This study belongs to the research program "Cooperation in Social and Economic Relations", which enjoys a waiver from Utrecht University's Institutional Review Board (IRB). Approval number: FETC17-028.

[*]Utrecht University School of Economics, Kriekenpitplein 21-22, 3584 EC Utrecht.

[†]Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081HV, Amsterdam & Tinbergen Institute & Radboud University, Institute for Management Research, Heyendaalseweg 141, 6525AJ Nijmegen.

[‡]Fernuniversität Hagen, Universitätsstr. 11, 58097 Hagen. Corresponding author: bastian.westbrock@fernuni-hagen.de

1 Introduction

We are involved in many social interactions in our daily lives and constantly struggle to divide our attention, time, effort, or resources between our friends, neighbors, and co-workers. Many experimental and empirical studies suggest that social preferences shape our behavior in such interactions. Yet, it is not clear how social preferences play out in a network of interdependent social interactions. How do individuals balance their selfish interests with a potential desire to achieve a fair(er) distribution across the many social interactions in which they are involved? Do social preferences help to coordinate behaviour when selfish incentives give rise to multiple equilibria? Do they help to overcome the inequality imposed by the macro-level network structure of personal connections?

In this paper, we study these questions in the context of the seminal local public goods game by Bramoullé and Kranton (2007). The game has much in common with the social dilemmas described above. Players share their investments (time, effort, resources) with their neighbors in a fixed network structure. Yet, players make only so many investments until they fill the gap between their personally desired level of the public good and the investments made in their neighborhood. This results in multiple equilibria that greatly differ in terms of overall welfare and the distribution of payoffs among players.

Our first contribution is that we introduce social preferences into this game and show that they limit the number of equilibria. Towards this end, we incorporate an n-player version of the Charness and Rabin (2002) social preference model into the game, which captures several types of social preferences that real people have been shown to care about.[1] We then characterize the equilibria with social preferences, what we refer to as the other-regarding equilibria (ORE). Our main results show that social preferences imply a very intuitive payoff ranking condition. Only those ORE are supported in many networks where payoffs are ordered according to the number of connections the players have. In other words, players in more central network positions earn more in equilibrium. And, if all players share the same connections, they earn the same.

Our payoff ranking condition is thus consistent with the social norm that managers earn more than their subordinates and professors more than their students. But it is also consistent with the norm that two flatmates should share the kitchen work equally, just as two coworkers should share the credit for a joint project.[2] However, the condition also implies that social

[1] See Sobel (2005) for an excellent earlier review of the experimental and empirical evidence and Bellemare, Kröger, and Van Soest (2008), Falk, Becker, Dohmen, Enke, Huffman, and Sunde (2018), and Kerschbamer and Müller (2020) for more recent evidence.

[2] Our payoff ranking condition is also consistent with earlier work in the social network literature. For example, it rationalizes the payoff-ranking assumptions of earlier network exchange theories (e.g., Cook and Emerson, 1978), and it enters many modern network formation models in a reduced form (e.g., the co-authorship model of Jackson and Wolin-

preferences do not necessarily entail more equitable or efficient outcomes in a network. Rather, they just reinforce the (in-)equality that is already inherent in the network structure.

Do social preferences impose a payoff ranking on all networks alike? Our second contribution shows that the answer is no. Social preferences facilitate equilibrium selection in small, tightly connected networks and in very centralized star-like structures. They fail to select, in contrast, in loosely-connected local interaction structures. More generally, we show that equilibrium selection through social preferences depends on two properties of the network structure: *size* and neighborhood *nestedness*.

While the impact of group size on coordination problems is well understood, the role of nestedness has not yet received much attention in the economics literature.[3] The neighborhoods of two players are nested if the neighborhood of one player is contained in the neighborhood of the other player who has a (weakly) larger number of connections. We show that two social players only agree on an unambiguous payoff ordering between them when their neighborhoods are nested. In other words, the central mechanism through which social preferences select equilibria hinges on a stronger requirement than just that one player has more connections than the other.

The role of nestedness becomes most apparent in our results for the circle network without any nested neighborhood. There, our theory predicts the co-existence of equal-split equilibria on one hand and specialized equilibria on the other, where some players free ride on the investments of their neighbors.

However, also for all the networks with nested neighborhoods, the exact organization of these neighborhoods matters if one takes into account that not all players share the same social preference. It is a well-known fact that people differ in their social preferences depending on, for instance, age, gender, and education (e.g., Bellemare, Kröger, and Van Soest, 2008; Falk, Becker, Dohmen, Enke, Huffman, and Sunde, 2018). Incorporating preference heterogeneity adds an additional dimension to our story because it implies that players might disagree about how payoffs should be ordered in a network. To capture this dimension, we model a game where each player's social preference is randomly determined at the start and players have incomplete information about other players' preferences. We then derive sufficient conditions regarding a combination of *compatible* social preferences that lead to an unambiguous payoff ordering between two players in a nested neighborhood of a network. How easily this payoff ordering

sky, 1996). Moreover, our equal-sharing prediction for a two-player dyadic interaction is reminiscent of the experimental evidence on 50:50 sharing in two-player games in the lab (Andreoni and Bernheim, 2009).

[3]Nestedness is a well-known topology of many ecological systems (Mariani, Ren, Bascompte, and Tessone, 2019) and many emergent social networks (König, Tessone, and Zenou, 2014; Belhaj, Bervoets, and Deroïan, 2016; Li, 2019; Olaizola and Valenciano, 2020). All the more surprising, our study is to the best of our knowledge the first to stress the functional importance of nested neighborhoods.

is reflected in the entire network structure not only depends on the network's degree of nestedness but also on how the nested neighborhoods are organized. Our results on the star network and the complete network most clearly illustrate that.

The second part of our paper validates the central mechanisms and predictions of our theory in an experiment. Our experiment has a number of design features to facilitate the test. First, subjects play the Bramoullé and Kranton (2007) game on a set of networks that differ widely in terms of nestedness. Second, we employ a large strategy space to allow for the full set of equilibria in the Bramoullé and Kranton (2007) game and for social comparison-motivated deviations thereof to emerge. Third, to ensure that subjects play equilibria at all, we let them play in continuous-time such as in Callander and Plott (2005), Berninghaus, Ehrhart, and Ott (2006), or Goyal, Rosenkranz, Weitzel, and Buskens (2017).

In the next section, we relate our contributions to the literature. Section 3 presents the game and our theoretical predictions. Section 4 describes the experiment. Sections 5–7 present the findings, and Section 8 discusses their implications. The proofs of all our statements, additional evidence from our experiment, and the replication instructions can be found in the appendix.

2 Related literature

Our study relates to the literature on social networks and on social preferences.[4] Our first contribution that social preferences facilitate equilibrium selection is related to one of the central topics in the recent network literature. As Charness, Feri, Meléndez-Jiménez, and Sutter (2014) put it:

> A critical problem for network theory is that even simple games have multiple equilibria, so that a great variety of outcomes are consistent with theoretical analysis. This naturally limits the predictive power of the theory and the scope of policy recommendations, since multiple equilibria make it difficult-to-impossible to offer definitive advice regarding how such labor markets, search markets, etc. should be organized. To make meaningful policy recommendations, it is crucial to determine which equilibrium is likely to occur (Charness, Feri, Meléndez-Jiménez, and Sutter, 2014, p. 1617).

The studies of Bramoullé, Kranton, and D'Amours (2014) and Allouch (2015) make clear when the problem of equilibrium multiplicity is most severe: in local interaction games where players' investments are strategic substitutes, hence just the class of games looked at in this study. The theoreti-

[4] A comprehensive overview of theoretical and experimental work on social and economic networks can be found in Bramoullé, Galeotti, and Rogers (2016) and Choi, Kariv, and Gallo (2016). For a comprehensive overview of the work on social preferences, see Sobel (2005).

cal literature has proposed several equilibrium refinement concepts so far. Bramoullé and Kranton (2007) study the equilibria that are *stable* with regard to Nash tâtonnement, and Boncinelli and Pin (2012) look at the stochastically stable equilibria. Galeotti, Goyal, Jackson, Vega-Redondo, and Yariv (2010), in turn, show that a limitation of the agents' information about the network structure can lead, probably paradoxically, to a sharp refinement of the equilibrium set.[5] All these concepts roughly select the same type of equilibria: the periphery-sponsorship equilibria where players with fewer connections contribute more to the local public good in their neighborhood than players with more connections. We develop a novel refinement concept that selects just the same type of equilibria in the class of asymmetric, star-like network structures. The new feature of our social preference theory is that it also provides a very natural, and as our experiment shows, empirically relevant refinement of the equilibria in symmetric network structures. These refined equilibria receive no support by the previous concepts.

The experimental literature on equilibrium selection has tested the role of *incomplete information* and Nash tâtonnement *stability*. In the most comprehensive study to date, Charness, Feri, Meléndez-Jiménez, and Sutter (2014) investigate the role of uncertainty about the network structure in a series of one-shot binary-choice games where the strategies of two network neighbors are either strategic substitutes or strategic complements. They conclude, however, that uncertainty does not facilitate coordination per se. Rather, the guiding principle to equilibrium selection in their strategic-substitutes games is *risk dominance*. In an experiment similar to ours on the original Bramoullé and Kranton (2007) game with the same large strategy space as in our experiment, but a non-continuous-time design, Rosenkranz and Weitzel (2012) compare the predictions of risk dominance with those of Nash tâtonnement stability and *quantal response* theory. Their findings provide partial support for all three theories, mainly because the rate of equilibrium play is so low that discrimination between the theories is difficult. Common to both experiments is that most of their evidence stems from asymmetric network structures where they find just the class of equilibria as predicted by all the theories, including ours, that is, periphery-sponsorship equilibria. Thus, in a sense, social preferences and fairness considerations have never been given a chance in these experiments; either because equal division of payoffs was ruled out by design or because coordination on any equilibrium at all was already so difficult that attempts to equalize payoffs were not discernible from the data. Our experimental design gives social preferences a chance to play a role in equilibrium selection.

There are a few other network theories with socially concerned play-

[5] One should not forget the literature on network formation at this point and one of its main findings that an expansion of the players' strategy sets to also include the selection or exclusion of partners can help to refine the equilibrium set of the game played on the network (e.g., Galeotti and Goyal, 2010; Goyal, Rosenkranz, Weitzel, and Buskens, 2017; Riedl and Ule, 2002).

ers, notably Ghiglino and Goyal (2010), Immorlica, Kranton, Manea, and Stoddard (2017), Bourlès, Bramoullé, and Perez-Richet (2017), and Richefort (2018). Different from our setting, they investigate environments without local strategic interactions so that coordination is not an issue. Social preferences merely "shift" the unique equilibrium points in these theories.[6] Another difference between these studies and ours is that they focus on one specific type of social preference, whereas we allow for the empirically more relevant case of heterogeneous preferences. The only other experimental study of social preferences in network games that we are aware of next to ours is Zhang (2018). The author compares the predictive power of two social preference types (altruism and inequity aversion) in two network structures (star and circle) and concludes that altruism is the better predictor. Yet, just as the above-mentioned studies, Zhang (2018) considers a game without any strategic interaction so that social preferences merely shift the observed investments in the experiment. By contrast, social preferences in our context can make the difference between an equilibrium with a center- and a periphery-sponsored public good and thus the difference between being the sole contributor and a free rider.

Our second contribution regarding the role of network *size* and network *nestedness* for equilibrium selection is closely related to the early literature on coordination games in symmetric local interaction structures (e.g., Ellison, 1993; Goyal and Janssen, 1997). A typical finding in this literature is that coordination is impeded by group *size* and is facilitated by a high level of *clustering* in each player's neighborhood. These predictions have been experimentally confirmed by, for example, Berninghaus, Ehrhart, and Keser (2002). The existing work on the link between network topology and coordination in more complex network structures is, to the best of our knowledge, exclusively experimental. Cassar (2007) and Charness, Feri, Meléndez-Jiménez, and Sutter (2014) show that the positive impact of *clustering* on coordination extends to various richer network structures. This conclusion is also confirmed by the findings in Rosenkranz and Weitzel (2012). They also find, however, that coordination is equally likely in highly *centralized* but otherwise non-clustered networks. Compared to these studies, our contribution is the theoretical underpinning and experimental support for a so-far overlooked property of the network structure: neighborhood *nestedness*.[7] In particular, our findings on the complimentary role of the degree of nestedness and the ideal organization of nested neighbor-

[6] To be precise, the general-equilibrium economy of Ghiglino and Goyal (2010) and the multiple public goods game of Richefort (2018) entail some form of strategic interaction. Yet, their interactions yield a unique equilibrium point with or without socially concerned players.

[7] Even though the level of clustering and the number of nested neighborhoods are potentially correlated in many network structures, the most conducive network for coordination in our context is the star network, that is, a network with zero clustering but a single player who nests the neighborhoods of all other players. In contrast, the fully clustered complete network is hardly any more conducive than a cirle network.

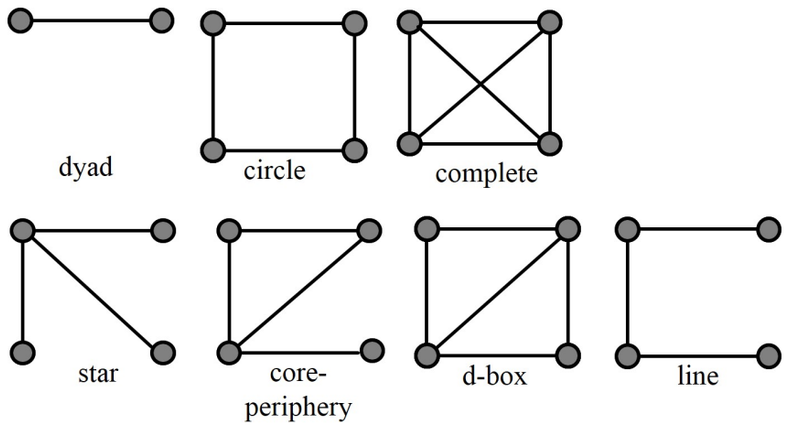


Figure 1: Networks in the experiment

hoods helps to explain the puzzle in Rosenkranz and Weitzel (2012). At least some degree of nestedness is important for coordination. But even more important is that a few players nest the neighborhoods of all others.

Finally, there is a link between our experimental findings and the large literature on social preferences.[8] The main contribution is that our network games fill the gap between the two extremes that are typically looked at in experiments: the complete interaction structure ($n : n$) and the star structure ($1 : n$), which are studied as stylized cases for market and bargaining interactions but which are less representative for many other social interactions.

3 Theory

3.1 Rules of the network game

We study the role of heterogeneous social preferences in the Bramoullé and Kranton (2007) public goods game. The rules are as follows: n players are embedded in a fixed network $g \in G$. Figure 1 illustrates some examples. All players simultaneously choose an investment that contributes to their own *local public good* and to that of their direct neighbors in g. Examples of such partner-independent investment are organizing parties for friends, project-specific investments by coworkers, experimentation with new tools, and neighborhood beautification expenses, all vis-à-vis the time or effort a person spends on her personal "projects".

Let $e_{-i} = \{e_1, e_2, ..., e_{i-1}, e_{i+1}, ..., e_n\}$ denote the contributions in all nodes of the network except in node i, and let $\mathcal{N}_i(g)$ denote the set of nodes in the

[8] Our findings are most closely related to those studies that go beyond pro-social behavior in small (1:1) interactions, notably the experiments on one-to-many bargaining (e.g., Roth, Prasnikar, Fujiwara, and Zamir, 1991; Schotter, Weiss, and Zapater, 1996), n-player public goods games with heterogeneous endowments (e.g., Fehr and Fischbacher, 2002; Buckley and Croson, 2006), and heterogeneous Cournot games (e.g., Maurice, Rouaix, and Willinger, 2013).

neighborhood of node i. The payoff of a player in node i is

$$\pi(e_i, e_{-i}, i) = b\big(e_i + e_{in}\big) - ce_i\,, \tag{1}$$

where $e_{in} \equiv \sum_{j \in \mathcal{N}_i(g)} e_j$ and where $b(\cdot)$ denotes the public-good benefit function, which satisfies $b'(0) > c > b'(\infty)$ and $b''(\cdot) < 0$, and c denotes a constant unit investment cost. In our experiment and at some points in our theory, we make use of the following linear-quadratic specification:

$$\pi(e_i, e_{-i}, i) = \begin{cases} \big(e_i + e_{in}\big)\big(\alpha - e_i - e_{in}\big) - ce_i & \text{if } e_i + e_{in} \leq \frac{\alpha - x}{2} \\ \frac{(\alpha - x)^2}{4} \;+\; x\big(e_i + e_{in}\big) - ce_i & \text{otherwise} \end{cases}, \tag{2}$$

where x is the minimal social benefit of an investment and α an intercept parameter.

A well-known prediction for the Bramoullé and Kranton (2007) game is that there exists a strictly positive investment level e^*, such that all players aim to fill the gap between e^* and the investments in their neighborhood (given that the latter do not already exceed e^*). In other words, the *payoff-maximizing* best-response function is given by

$$f(e_{-i}, i) = \begin{cases} e^* - e_{in} & \text{if } e_{in} \leq e^* \\ 0 & \text{otherwise} \end{cases}. \tag{3}$$

Moreover, there are multiple equilibria for every network structure. In the dyad in Figure 1, for example, or its n-player extension, the complete network, any investment profile is a Nash equilibrium as long as the sum of investments is equal to e^*. Similarly, in the star network, there are two Nash equilibria, one equilibrium where the center player provides the desired level e^* and another equilibrium where the periphery players invest each e^*.[9]

It is obvious that all these equilibria differ markedly in terms of the overall welfare and the payoff distribution they induce. In the context of coworker teams, for example, the prediction for a star-like team structure is that either the team leader does all the work or her subordinates do. Similarly, one prediction for a non-hierarchical team is that the workload is split equally among the team members; another prediction is that one coworker does all the work. This calls for reasonable equilibrium refinement concepts and experiments testing their importance. As we will see, social preferences serve both purposes; they can lead to a fine-grained selection in the equilibrium set and the equilibria they select are empirically relevant.

[9] More generally, in any network structure there exists a class of *specialized* equilibria, where some players contribute the desired investment level e^*, while their neighbors exert no effort, and a class of *distributed* equilibria, where every player exerts some effort (Bramoullé and Kranton, 2007).

3.2 A social preference function for network games

We first present our social preference model. The theoretical literature on social preferences has produced various meaningful preference models for two-player games or n-player symmetric games (see Sobel, 2005, for a review). Games on more complex interaction structures have been left out of the perspective, however.[10] Our model is an n-player extension of the distributional preference models of Charness and Rabin (2002) and Schulz and May (1989), which nests various distributional preferences such as altruism, inequity aversion, competitiveness, and spite.[11] It formulates a player's preferences in the following way:

$$U_s(e_i, e_{-i}, i) \;=\; \pi_i \;+\; \frac{1}{|R_s|} \sum_{j \in R_s} \left(\rho_s \, r_{ij} + \sigma_s \, s_{ij}\right) \pi_j \,, \tag{4}$$

where R_s denotes the player's reference group that satisfies $\mathcal{N}_i(g) \subseteq R_s \subseteq \mathcal{N} \backslash \{i\}$, ρ_s and σ_s denote two real-numbered preference parameters, $\rho_s \geq \sigma_s$, and

$$\begin{aligned} r_{ij} &= 1 \quad \text{if } \pi_i > \pi_j \text{ and } r_{ij} = 0 \text{ otherwise,} \\ s_{ij} &= 1 \quad \text{if } \pi_i < \pi_j \text{ and } s_{ij} = 0 \text{ otherwise.} \end{aligned}$$

This formulation says that utility is a linear combination of material payoffs and a social preference component. The latter captures the (dis-)utility players derive from comparing their payoffs with those of other players in the game. With whom a player compares is defined by the reference set R_s. The set might comprise any number of players in a network. It seems natural, however, that players only compare with their neighbors who they can directly influence, $R_s = \mathcal{N}_i(g)$, or benchmark their payoffs against everyone else in a network, $R_s = \mathcal{N} \backslash \{i\}$.

Players thereby distinguish between peers in their reference group who are behind ($\pi_i > \pi_j$) and peers who are ahead ($\pi_i < \pi_j$). The parameters ρ_s and σ_s then govern the (dis-)utility players derive from comparing with those behind and those ahead. Depending on the specific parameter combination, the model describes various meaningful preference types. Unconditional *altruists* ($\rho_s \geq \sigma_s > 0$), for example, are always willing to give up some of their own payoffs to help others. *Social-welfare* types, in contrast, withdraw their assistance when they are behind everyone else ($\rho_s > \sigma_s = 0$). In the negative domain, *spiteful* types ($0 > \rho_s \geq \sigma_s$) are always willing to forego some of their own payoffs to lower the payoffs of others, while

[10] The exception here is the studies on social networks mentioned in the literature review that have developed their own interdependent preference models. Utility model (4) nests several of these as special cases. In particular, Ghiglino and Goyal (2010) consider what we define as *spitefulness*, and Immorlica, Kranton, Manea, and Stoddard (2017) consider *competitiveness*. The model in Bourlès, Bramoullé, and Perez-Richet (2017), in contrast, describes a situation where players know each other well and, accordingly, include each others' utilities rather than payoffs in their own utility functions.

[11] For a recent experimental test of this model, see Bruhin, Fehr, and Schunk (2019).

competitive types ($0 = \rho_s > \sigma_s$) refrain from these welfare-reducing actions when they are ahead of everyone else. The two domains are connected by the *inequity-averse* types ($\rho_s > 0 > \sigma_s$) who forfeit some of their payoffs to players who are behind but who are willing to incur welfare-reducing actions to lower the payoffs of those who are ahead.

Model (4) thus captures a wide range of empirically relevant preference types and, as we will see, is nevertheless simple enough to produce sharp predictions in the context of the Bramoullé and Kranton (2007) game.[12]

3.3 Other-regarding equilibria: general results

Equipped with this utility model, we now turn to the characterization of the equilibria of the extended Bramoullé and Kranton (2007) game, henceforth the ORE. We think of a game where the parameters of utility model (4) are randomly determined for each player before the start of the game. Specifically, the type $t_s \equiv (\rho_s, \sigma_s, R_s)$ of a player in node i is determined by an i.i.d. draw from a probability distribution over the support $\mathcal{T}_i = \{\tau_1, \tau_2, ..., \tau_t\}$, where $\mathcal{T}_i$ defines a *finite* subset of the set of all feasible combinations of τ_s. The actual combination of player types in a game, $\omega \equiv (t_s 1, t_{s'} 2, ..., t_{s''} n)$, is then an i.i.d. random variable drawn from the set $\Omega \equiv \mathcal{T}_1 \times ... \times \mathcal{T}_n$. Moreover, an ORE is a Bayesian Nash equilibrium of a game with incomplete information about each player's type, where the strategy of a player is a mapping $\Sigma_i : \mathcal{T}_i \to \mathbb{R}_+$.

Let $f^{or}(t_s, i, e_{-i})$ denote the best-response investment of a type-t_s player in node i against the "expected" investments in the other nodes:

$$e_{-i} \equiv \left(e_{\tau_1 1}, ..., e_{\tau_t 1}, ..., e_{\tau_1 i-1}, ..., e_{\tau_t i-1}, e_{\tau_1 i+1}, ..., e_{\tau_t i+1}, ..., e_{\tau_1 n}, ..., e_{\tau_t n}\right).$$

Point predictions for this best-response investment are difficult to make. The reason is that the optimal response of a player not only depends on her social preference type but also on her relative standing vis-à-vis every single other player (τ_s, i) in her reference group. Nevertheless, we can define several general conditions that a best response, and an ORE, must satisfy.

First, players do not deviate "too much" from a payoff-maximizing best response. Instead, a player's type t_s defines how far away her best-response investment is from a pure payoff-maximizing investment. To measure this deviation, we define

$$f^{non}\big((0,0), i\big) \equiv f^{non}\big((\rho_s = 0, \sigma_s = 0), i, e_{-i}\big)$$

as the unconstrained (thus possibly negative) optimal response of a payoff

[12] In fact, model (4) circumvents an ambiguity of the original Charness and Rabin (2002) model in the context of the Bramoullé and Kranton (2007) game. Apart from the n-player extension, another major difference between the two models is that the absolute *level* of other players' payoffs enters utility (4), rather than their *relative* payoff vis-à-vis the focal player. This modification avoids a counter-intuitive prediction of the original model, which we address more explicitly in our robustness analysis in Experimental Appendix B.3.

maximizer that would make the investment in a player's stead. When payoffs are given by the linear-quadratic function (2), for example, then this optimal response is simply given by

$$f^{non}\big((0,0),i\big) \;\; = \;\; e^* - E[e_{in}]\,,$$

where $\mathbb{E}[e_{in}]$ denotes the expected investments of i's neighbors.

We then define the scalar $\epsilon_s \in \mathbb{R}_+$ that measures the maximal absolute deviation that a type t_s would be willing to make for the sake of a fairer outcome. It then follows that $f^{or}(t_s, i, e_{-i})$ is constrained by

$$\max\big\{f^{non}\big((0,0),i\big) - \epsilon_s\,;\,0\big\} \;\leq\; f^{or}\big(t_s,i,,e_{-i}\big) \;\leq\; f^{non}\big((0,0),i\big) + \epsilon_s. \quad (5)$$

In other words, ϵ_s is a measure of the *strength* of a player's social preferences and

$$\epsilon \equiv \max\{\epsilon_\tau \,|\, \tau \in \mathcal{T}_i, i \in \mathcal{N}\}$$

measures the maximal preference strength of a player group. The following result shows how to map a player's t_s into a value for ϵ_s.

Lemma 1. *Suppose that players' utilities are defined by payoff function (1) and the social preference model (4) with parameters $t_s = (\rho_s, \sigma_s, R_s)$. A player's social preference strength, ϵ_s, is given by*

$$\begin{aligned}
&\textit{for an altruist or a social-welfare type } (\rho_s \geq \sigma_s \geq 0): \;\; \epsilon_s^p \\
&\textit{for an inequity-averse type } (\rho_s > 0 > \sigma_s > -1): \;\; \max\{\epsilon_s^p\,;\,\epsilon_s^n\} \\
&\textit{for a competitive or spiteful type } (0 \geq \rho_s \geq \sigma_s > -1) \;: \;\; \epsilon_s^n
\end{aligned}$$

where

$$\begin{aligned}
\epsilon_s^p \;\; &= \;\; (b')^{-1}\big(\frac{c}{1+\rho_s}\big) - e^* \\
\epsilon_s^n \;\; &= \;\; e^* - \max\big\{(b')^{-1}\big(\frac{c}{1+\sigma_s}\big)\,;\,0\big\}.
\end{aligned} \quad (6)$$

See Theoretical Appendix A.1 for the proof. The expressions in (12) are intuitive. Because the optimal investment of a payoff maximizer, e^*, is defined by $e^* \equiv (b')^{-1}(c)$, the expressions suggest that social preferences do nothing but alter the personal unit cost of a public goods investment. Competitive and spiteful players choose an investment as if they had a higher cost, altruists and social-welfare types invest as if their cost was lower.

Thus, at first sight, social preferences do *not* seem to facilitate equilibrium selection. Quite on the contrary, the expressions in (5) and (12) suggest that the set of ORE is wider than the set of payoff-maximizing equilibria because additional equilibria can be maintained where some or all players deviate from a payoff-maximizing investment. Yet, other-regarding players do not deviate from a payoff maximum in an arbitrary way. As they ultimately strive for a certain payoff ordering, the set of ORE profiles is con-

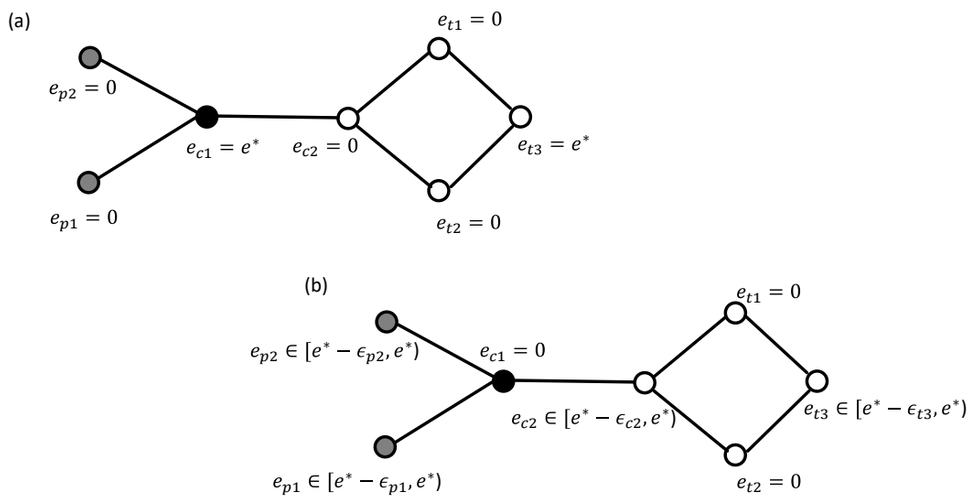


Figure 2: Payoff-maximizing and refined other-regarding equilibria

NOTES: Panel (a) shows one of three specialized payoff-maximizing equilibrium profiles that is not a *refined* ORE. Panel (b) exhibits a *refined* ORE, which coincides with one of the other two specialized payoff-maximizing equilibrium profiles when $\epsilon \equiv \max\{\epsilon_\tau \mid \tau \in \mathcal{T}_i, i \in \mathcal{N}\} \to 0$. The gray and black nodes indicate players in nested neighborhoods.

strained in a systematic way. To see how the equilibrium set is constrained, let us look at the following example.

Example 1. Consider the investment profile in Figure 2 Panel (a). Suppose that all players are inequity-averse and suppose they include only their direct neighbors in their reference group, both of which is common knowledge for all players. In that case, the profile cannot be maintained in an ORE, despite being a payoff-maximizing equilibrium. This is because player $c1$ (and player $t3$) would necessarily want to reduce their investments below e^* as they are the only ones who contribute in their neighborhood and therefore feel exploited. At the same time, their neighbors would necessarily want to increase their investments (from zero) because they feel guilty. The profile in Panel (a) can thus be ruled out as an ORE.

To avoid any such deviations, we need to look for a profile where player $c1$ earns more than at least one of her neighbors, say player $p1$. This is because player $c1$'s envy of players $p2$ or $c2$ may then be balanced out against her guilt towards $p1$. Such a profile is displayed in Panel (b). In fact, that profile can be maintained in an ORE, despite the fact that the other central player $c2$ is one of the players who earns the least.

Why are the central players $c1$ and $c2$ treated differently in this example? The reason is that the *payoff ordering* in player $c1$'s neighborhood is tied to the fact that the neighborhood of $c1$ *nests* the neighborhoods of players $p1$ and $p2$. This nestedness, in turn, implies that players $p1$ and $p2$ do not receive any investments that $c1$ does not have access to. Combined with their aversion to inequity, players $p1$ and $p2$ thus cannot earn more than player $c1$ because their feelings of guilt would make them increase their investments to help $c1$. The same cannot be said about player $c2$ and her neighbors,

however. All three of them have access to at least one other player, who $c2$ does not have access to and who contributes to their local public good. And, because the total investments that $c2$'s neighbors receive in Panel (b) are far beyond their personally desired level of the public good (when their ϵ_s is sufficiently), they are not willing to make the extra investment that would be needed for a more equitable outcome for player $c2$. The profile in Panel (b) can thus be maintained in an ORE.

Equilibrium selection through social preferences is tied to an additional condition, however. To see which, consider the profile in Panel (a) again.

Example 2. Suppose that, instead of all players being inequity-averse, player $c1$ is of a social-welfare type while players $p1$, $p2$, and $c2$ are competitive or spiteful. The profile in Panel (a) can then be maintained in an ORE—in addition to the ORE in Panel (b)—because $c1$'s neighbors do not feel guilty any longer (maintain $e_j = 0$), while $c1$ looks after herself (plays $e_{c1} = e^*$).

Why can the profile in Panel (a) been ruled out as an ORE when all players are inequity-averse, but not when they have the preference constellation of Example 2? The reason is that the preferences of the players must be *compatible*, in the sense that players must prefer a payoff ordering that "fits" the position they occupy in a network. As we have seen above, competitive or spiteful types in the central positions of a network and social-welfare or altruistic types in the peripheral positions do *not* meet this requirement. But a group of only inequity-averse types does meet them.

Generalizing from here, we define a *refined ORE* as an other-regarding equilibrium profile where at least some (or all) players' social preferences are *compatible* so that their payoffs must be ordered in an arbitrary network g. The following result derives the general conditions under which this payoff-ordering property must be satisfied.

Lemma 2. *Consider two neighbors i and j in a nested neighborhood in a network g, that is, $\mathcal{N}_j(g) \cup \{j\} \subseteq \mathcal{N}_i(g) \cup \{i\}$. Suppose that i's and j's social preferences are compatible and that this is common knowledge. That is, their type sets $\mathcal{T}_i = \mathcal{T}_i^*$ and $\mathcal{T}_j = \mathcal{T}_j^*$ satisfy*

$$\begin{gathered} \mathcal{T}_i^* \equiv \{\textit{comp., spite, inequity av.}\} \ \wedge \ \mathcal{T}_j^* \equiv \mathcal{T} \backslash \{\textit{spite}\} \\ \vee \\ \mathcal{T}_i^* \equiv \mathcal{T} \backslash \{\textit{altruist}\} \quad \wedge \quad \mathcal{T}_j^* \equiv \{\textit{social welfare, altruist, inequity av.}\}. \end{gathered} \tag{7}$$

In a refined ORE, it must be that $\pi(i,\omega) \geq \min_{k \in R_i}\{\pi(k,\omega)\}$ OR $\pi(j,\omega) \leq \max_{k \in R_j}\{\pi(k,\omega)\}$ for at least one $\omega \in \Omega^* \equiv \mathcal{T}_1 \times ... \mathcal{T}_i^* \times ... \mathcal{T}_j^* \times ... \mathcal{T}_n$.

The proof is as follows. Suppose that player j's neighborhood is nested in player i's, and suppose their preferences are compatible but that, contrary to the statement, all types of player i (j) earn strictly less (more) than all players in their respective reference groups, that is, $\pi(i,\tau) < \min_{k \in R_\tau}\{\pi(k,\omega)\}$

AND $\pi(j,\tau') > \max_{k\in R_{\tau'}}\{\pi(k,\omega)\}$ for all $\omega \in \Omega^*$. Then, it must hold that[13]

$$0 \;<\; e_{\tau i} \;\leq\; f^{non}\big((0,0),i\big) \;<\; f^{non}\big((0,0),j\big) \;\leq\; e_{\tau' j}\,, \tag{8}$$

because either player i feels exploited or player j feels guilty (or both).

However, this implies for the type t_s with the highest investment among all $\tau \in \mathcal{T}_i^*$ and the type t_t with the lowest investment among all $\tau' \in \mathcal{T}_j^*$ (for which $e_{t_s i} < e_{t_t j}$ still holds) that $\mathbb{E}\pi(t_s,i) > \mathbb{E}\pi(t_t,j)$. This is because t_s receives, in expectation, a larger local public good than t_t and pays a lower cost. Hence, we arrive at a contradiction to the supposed payoff ordering. ■

The intuition extends immediately from Examples 1 and 2. The fact that player j feels obliged to increase her investment beyond the payoff-maximizing level to "help player i out" makes it impossible that all types of i earn the least in their reference group and all types of j earn the most. In a refined ORE, the payoffs of i and j must, therefore, be ranked according to the ordering property in Lemma 2. There are several remarks to be made about this property:

1) The required condition on the preference constellation of players i and j is, for example, satisfied when all types of i and j are either competitive, social welfare-concerned, or inequity-averse.

2) The assumption that the types of players i and j are drawn from the restricted sets $\mathcal{T}_i^*$ and $\mathcal{T}_j^*$ is crucial. Otherwise, there could be a type of player j who is *not* willing to help player i and who thus earns more than everybody else in her reference group (even though the actual t_t is not that type). The assumption of common knowledge is also important. Otherwise, a player j of the correct type $t_t \in \mathcal{T}_j^*$ might still mistakenly believe that player i is not needy or player i might believe that j is not willing help her out, etc. In other words, it is immaterial for Lemma 2 which social preference types players i and j exactly have as long as their types stem from the compatible sets $\mathcal{T}_i^*$ and $\mathcal{T}_j^*$, and this is common knowledge.

3) When the exact type of each player is commonly known, Lemma 2 can be significantly strengthened. This is because the state set Ω^* reduces

[13] To see why $f^{non}((0,0),i) < f^{non}((0,0),j)$, note that $f^{non}((0,0),k)$ for $k \in \{i,j\}$ is given by the solution to the first-order condition

$$\frac{\partial \mathbb{E}\pi}{\partial e_k}\big(k, f^{non}((0,0),k), e_{-k}\big) \;=\; \sum_{\omega\in\Omega} p(\omega)\, b'\big(f^{non}((0,0),k) + \sum_{l\in\mathcal{N}_k(g)} e_{\tau l}\big) - c \;=\; 0\,.$$

Now, because (i) the neighborhood of j is nested in the neighborhood of i, it follows that $\sum_{k\in\mathcal{N}_i(g)\setminus\{j\}} e_{\tau k} \geq \sum_{k\in\mathcal{N}_j(g)\setminus\{i\}} e_{\tau k}$. Moreover, because (ii) $e_{\tau' j} \geq f^{non}\big((0,0),j\big)$ for all $\tau' \in \mathcal{T}_j^*$ while $e_{\tau i} \leq f^{non}\big((0,0),i\big)$ for all $\tau \in \mathcal{T}_i^*$ (with one inequality being strict), it follows that the first-order condition of j is binding at a larger value of $f^{non}((0,0),j)$ than the first-order condition of i.

to a singleton in this case. Straightforward application of the conditions in Lemma 2 implies that either player i must earn more than everybody else in her reference group or player j must earn less. That is, it must holds that

$$\pi(i,t_i) \geq \min_{k\in R_{t_i}}\{\pi(k,t_k)\} \quad \text{OR} \quad \pi(j,t_j) \leq \max_{k\in R_{t_j}}\{\pi(k,t_k)\}$$

for all $k \in R_{t_i}$ and $k \in R_{t_j}$. The same can be said about a dynamic extension of our incomplete information game—like the extension implemented in our experiment—where the game is repeated over T round and players observe in every round t all the investments of the other players. A myopic best-response dynamic will lead to an ORE in which the payoffs of players i and j are ordered according to Lemma 2. All that is required in this case is that the *actual* types of i and j are compatible.

4) The payoff-ranking property can also be significantly strengthened when assuming a narrower preference constellation or a narrower class of networks. If $\mathcal{T}_i = \mathcal{T}_i^{**}$ and $\mathcal{T}_j = \mathcal{T}_i^{**}$, for example, where

$$\mathcal{T}_i^{**} \equiv \{\text{comp., spite}\} \ \wedge \ \mathcal{T}_j^{**} \equiv \{\text{payoff max., social welfare, altruist}\}, \quad (9)$$

player j must earn weakly less than player i, that is, $\pi(j,\omega) \leq \pi(i,\omega)$ in at least one state $\omega \in \Omega^{**} \equiv \mathcal{T}_1 \times ...\mathcal{T}_i^{**} \times ...\mathcal{T}_j^{**} \times ...\mathcal{T}_n$. If player i is of type $\tau \in \mathcal{T}_i^{**}$ and is connected to all other players j in the network, who have compatible preferences (i.e., $\tau' \in \mathcal{T}_j^{**}$), it even holds that $\pi(i,\omega) \geq \max_{j\neq i}\{\pi(j,\omega)\}$ for at least one ω. The same can be said when only a single one of player i's neighbors has compatible preferences. What is needed in this case, instead, is that players include only their direct neighbors in their reference group, that is, $R_\tau = \mathcal{N}_k(g)$ for all players k and their types τ.

Even more can be said when we assume that the social preferences of all players are sufficiently "small" in addition and assume that players are paid according to the linear-quadratic function (2). Our first result shows that, as expected, the set of ORE converges to the set of money-maximizing equilibria in that case.

Proposition 1. *Suppose that payoffs are given by the linear-quadratic function (2). When $\epsilon \to 0$, the set of ORE coincides with the set of money-maximizing equilibria.*

The proof is simple. In an ORE, it holds for all active players with $e_{\tau i} > 0$ that the total investment in their neighborhood, $e_{\tau in} \equiv e_{\tau i} + \mathbb{E}[e_{in}]$, is constrained by

$$f\big((0,0),i\big) + \epsilon + \mathbb{E}[e_{in}] \ \geq \ e_{\tau in} \ \geq \ f\big((0,0),i\big) - \epsilon + \mathbb{E}[e_{in}]\,,$$

where the boundaries follow from condition (5). Using the best-response condition of a payoff maximizer who makes the decision in their position, $f\big((0,0),i\big) = e^* - \mathbb{E}[e_{in}]$, this simplifies to

$$e^* + \epsilon \geq e_{\tau in} \geq e^* - \epsilon.$$

Thus, in the limit of $\epsilon \to 0$, we obtain the necessary and sufficient equilibrium condition for an active payoff maximizer, $e_{in} = e^*$. Similar, the total investment received by a passive type with $e_{\tau i} = 0$ must satisfy $e_{\tau in} \geq e^* - \epsilon$ because a payoff maximizer in position i would just be indifferent between contributing and not when she receives e^* from her neighbors and so there are social types that are just satisfied with $e^* - \epsilon$. Thus, again, we obtain the equilibrium condition of a passive payoff maximizer, $e_{\tau in} \geq e^*$, in the limit.

Together, this also implies that all types in the same position must invest the same. The reason is that there cannot be two types τ and τ' in position i for which $e_{\tau i} > e_{\tau' i} \geq 0$, $\lim_{\epsilon \to 0} e_{\tau in} = e^*$, and $\lim_{\epsilon \to 0} e_{\tau' in} \geq e^*$ simultaneously hold. ■

Together, Lemma 2 and Proposition 1 have some strong implications for the equilibrium investments in a network. Remember that our social preference theory does not help to refine the set of equilibria in a non-nested network structure, such as the circle network of Figure 1. However, when a network has at least one nested neighborhood and players in this neighborhood have compatible social preferences, then refined ORE even become a proper subset of the money-maximizing equilibria. Our next result characterizes the set of refined ORE for a group of players who mutually nest each others' neighborhoods and who only include these neighbors in their reference group.

Proposition 2. *Consider a network g with a fully interconnected, but otherwise isolated, component $\mathcal{C}(g')$, with $g' \subseteq g$ such that $\forall i, j \in \mathcal{C}(g')$ and $k \in \mathcal{N} \backslash \mathcal{C}(g')$, $g_{ij} = 1$ and $g_{ik} = 0$. Suppose that all players $i \in \mathcal{C}(g')$ only compare with their direct neighbors (i.e., $R_i = \mathcal{C}(g') \backslash \{i\}$) and that $\epsilon \to 0$. In a refined ORE, it must be*

$$e_i = e_j = \frac{e^*}{|\mathcal{C}(g')|} \quad \textit{for all } i, j \in \mathcal{C}(g').$$

The result, which is proven in Theoretical Appendix X, shows that social preferences lead to a very fine-grained selection in the set of money-maximizing equilibria when they are compatible. Remember that in a fully interconnected component, any investment profile can be a money-maximizing equilibrium as long as $\sum_{i \in \mathcal{C}(g')} e_i = e^*$. These equilibria cannot be refined by means of several established concepts, such as Nash tâtonnement stability, efficiency, or stochastic stability. In a refined ORE, in contrast, all players make the exact same investment. The intuition extends immediately from

Lemma 2. Suppose that the players $i \in \mathcal{C}(g')$ have compatible preferences,[14] but that contrary to the statement not all their investments are equal. The fact that their neighborhoods are mutually nested means that there is at least one player who earns (weakly) less than everybody else, and another player who earns (weakly) more. At least one of them would feel insulted in her understanding of fairness and adjust her strategy. Such an adjustment can only be avoided when all players make the exact same investment.

We now turn to our equilibrium characterization for a general network structure g. Here, we can say more when we focus on the set of *specialized* investment profiles. Bramoullé and Kranton (2007, p. 482) define a specialized profile as a profile where every player either exerts the maximum amount of effort $e_i = e^*$ (active) or exerts no effort $e_i = 0$ (inactive). The authors then show that a specialized profile is a money-maximizing equilibrium if and only if its set of of specialists is a *maximal independent set* of a network g (Bramoullé and Kranton, 2007, Theorem 1).[15] Social preferences refine the equilibrium set in the following way.

Proposition 3. *Suppose that all players compare with their direct neighbors (i.e., $R_i = \mathcal{N}_i(g)$) and that $\epsilon \to 0$. If a specialized profile is a refined ORE, then its set of active players is a maximal independent set I of network g with the property that every player i who nests another player's neighborhood in g is inactive (i.e., $i \notin I$).*

The first part that a specialized profile is a refined ORE only if its active players form a maximial independent set I of g follows immediately from Proposition 1 and Bramoullé and Kranton (2007, Theorem 1). The second part that a nesting player must be inactive follows by contradiction. In particular, suppose to the contrary that player i nests another player j's neighborhood but that i is active. It must then be $e_j = 0$. In fact, it must be $e_k = 0$ for all $k \in \mathcal{N}_j(g)$ since i nests j's neighborhood. Hence, player j earns the most in her reference group and player i earns the least where in fact $\pi(i) < \pi(j)$. This is however a contradiction to Lemma 2 according to which the payoffs of players i and j with compatible social preferences must be conversely ordered. ■

For example, applied to the star, the line, and the d-box networks in Figure 1, the result implies that the public good must be sponsored entirely by the peripheral players.[16] Intuitively, the reason is that the requirement

[14] This means that for all $i \in \mathcal{C}(g')$ (a) no $\mathcal{T}_i^c$ contains a type of the set {altruist, spite}, (b) no two or more $\mathcal{T}_i^c$ contain a *money maximizer*, and (c) no two or more $\mathcal{T}_i^c$ contain a distinct type of the set {money max., social welfare, comp.}.

[15] The graph-theoretic concept of a *maximal independent set* is defined as follows. An independent set I of a graph g is a set of nodes such that no two nodes that belong to I are linked; i.e., $\forall i, j \in I$, $g_{ij} = 0$. An independent set I is maximal when it is not a proper subset of any other independent set. Moreover, it satisfies the property that every node either belongs to I or is connected to a node that belongs to I.

[16] In contrast, no specialized profile can be a refined ORE in the core-periphery network because Proposition 1 would clash with the requirement that the duo players want to receive at least e^*. As we will see in our experimental predictions, the core-periphery network has a non-specialized ORE though.

of Lemma 2 that the center player must earn more than at least one other peripheral player can only be met when the investments are made in the periphery. Hence, the payoff ranking of Lemma 2 even translates into an investment ranking so that players who nest other players' neighborhoods will also invest less. We will make extensive use of this refinement of the equilibrium set in our experimental predictions.

4 Experiment

Broadly speaking, our theory predicted that (i) compatible social preferences facilitate coordination in a local interaction game with multiple equilibria, but (ii) whether coordination succeeds also depends on the nestedness of the network structure. In the following, we test these predictions in an experimental implementation of the Bramoullé and Kranton (2007) game.

4.1 Experimental game

We administered a dynamic extension of the Bramoullé and Kranton (2007) game on the whole set of two- and four-player networks shown in Figure 1. These networks are ideal for our theory testing because they differ widely in terms of their degree of nestedness and their organization of nested neighborhoods. The circle network, for example, is a network without any nested neighborhoods, while the dyad and complete networks only consist of mutually nesting neighborhoods. In the set of asymmetric networks, in turn, we have two networks with a single player in the central nesting position (star, core periphery) and two flatter hierarchies with two nesting centers (line, d-box).

Our experimental games differ from the original static game in Bramoullé and Kranton (2007) because experience with earlier experiments on this game made it clear that subjects find it difficult to coordinate their choices. Coordination was particularly difficult in versions of the game that adopted the original large strategy space (e.g., Rosenkranz and Weitzel, 2012). However, as at least some equilibrium play is essential for our theory testing, we administered a dynamic extension that nevertheless retains some key properties of the original game.

Specifically, following Callander and Plott (2005) and Berninghaus, Ehrhart, and Ott (2006), every experimental game lasted between 30 and 90 seconds. The final decision moment, t^{max}, was randomly determined during a game by a draw from the uniform distribution on $[30, 90]$. Starting from a situation of zero investments, subjects could continuously update their investments, choosing from the entire set of positive integer values. Full information about the momentary investments of all other players was continuously provided and updated five times per second. Moreover, information

about the momentary payoffs was indicated by the size of each player's node on the screen (see the screenshot in Appendix C.2). Nevertheless, the actual payoff of a game was solely determined by the momentary investments at t^{max}. All earlier decisions were payoff-irrelevant. Specifically, subjects were rewarded based on the linear-quadratic payoff function (2) with $x = 1$, $\alpha = 29$, and $c = 5$. This means that the individual payoff-maximizing contribution level is given by $e^* = 12$ and that after a total investment of 14 units in a player's neighborhood, payoffs change linearly.

Coordination is facilitated in our dynamic extension of the original game for at least two reasons. First, subjects can observe the entire history of play, allowing them to learn about the preferences and motives of the other players. Second, subjects can observe all momentary investments, which is all a player with a distributional social preference needs to make a "fair" best response. At the same time, the implemented random stopping rule avoids last round effects.[17]

4.2 Predictions and hypotheses

In the following, we fully characterize the sets of *refined* ORE for the seven networks in the experiment, whereby we look at the empirically relevant case of significant deviations from a pure money-maximizing best response (i.e., large values for ϵ). The proofs of all our statements can be found in Theoretical Appendix A.3.[18]

Star, core periphery, and d-box: Remember that two markedly different investment profiles can be supported in a Nash equilibrium in these networks when players are pure payoff maximizers: one *periphery-sponsorship* equilibrium where the center player(s) (denoted as $i \in \mathcal{C}$) free ride(s) on all the other players, who each contribute $e_j = e^*$, and another *center-sponsorship* equilibrium where the other players, in particular the peripheral players (denoted as $j \in \mathcal{P}$), free ride on the center(s).

When players have social preferences and these preferences are compatible, Lemma 2 predicts that the center player earns weakly more than at least one other player,

$$\pi(i, \omega) \geq \min_{j \neq i} \{\pi(j, \omega)\} \quad \text{for all } i \in \mathcal{C}, j \in \mathcal{N} \backslash \mathcal{C},\ \omega \in \Omega^*.$$

The combination of compatible preferences and limited preference strength

[17] In fact, as shown in Section 5, we observe a strong increase in the frequency of (static) equilibria being played compared with, for example, Rosenkranz and Weitzel (2012). Moreover, although our setup may make it likely that subjects collude at the beginning of a game to reach Pareto-superior outcomes in its continuation, we find no indication of collusion.

[18] The proofs contain the equilibrium characterizations for both the incomplete-information game of Section 3 and the perfect-information game implemented in the experiment. For an easier link with our empirical analysis, we present below the more fine-grained predictions for the perfect-information game.

leads to another powerful refinement of this conditions. The reason is that the above payoff-ordering property can only be guaranteed for when the public goods is entirely provided by the non-center players. To formalize this, suppose that players' social preferences are compatible and small, that is, suppose that $\omega \in \Omega^h \equiv \Omega^{h_1} \cup \Omega^{h_2}$ with $h \in \{star, core, dbox\}$ and

$$\Omega^{h_1}: \ \mathcal{T}_i = \mathcal{T}_i^* \text{ for all } i \in \mathcal{C}, \ \mathcal{T}_j = \mathcal{T}_j^* \text{ for all } j \in \mathcal{N}\backslash\mathcal{C}, \text{ and } \epsilon < \overline{\epsilon}^h,$$

$$\Omega^{h_2}: \ \mathcal{T}_i = \mathcal{T}_i^* \text{ for all } i \in \mathcal{C}, \ \mathcal{T}_j = \mathcal{T}_j^* \text{ for at least one } j \in \mathcal{P}, \ R_\tau = \mathcal{N}_k(g) \text{ for all } \tau \in \mathcal{T}_k, k \in \mathcal{N}, \text{ and } \epsilon < \overline{\epsilon}^h,$$

where $\mathcal{T}_i^*$ and $\mathcal{T}_j^*$ are defined in Lemma 2 and $\overline{\epsilon}^{star} = \overline{\epsilon}^{core} > \overline{\epsilon}^{dbox}$ are defined in the proof. Then, all refined ORE entail a periphery-sponsored public good where

$$e_{\tau i} = 0 \quad \text{and} \quad e^* - \epsilon \leq e_{\tau j} \leq e^* + \epsilon \quad \text{for all } i \in \mathcal{C}, j \in \mathcal{P}, \ \omega \in \Omega^h.$$

Hence, social preferences select the "natural" equilibrium in the star, the core periphery, and the d-box, where the center player(s) earn strictly more than *every* other player.

Dyad and complete networks: Suppose that all players' social preferences are compatible. That is, suppose that $\omega \in \Omega^c = \mathcal{T}_1^c \times ... \times \mathcal{T}_n^c$ where $c \in \{dyad, comp\}$ and where the conditions for players i and j in Lemma 2 mutually apply to every pair of nodes. In a refined ORE, it must be

$$e_{\tau i} = e_{\tau' j} = e \quad \text{for all } \omega \in \Omega^c \text{ , where } \frac{e^* - \epsilon}{n} \leq e \leq \frac{e^* + \epsilon}{n}.$$

Hence, our experimental predictions extend the result in Proposition 2 in that social preferences also select an equal-split equilibrium when players have significant social preferences. As we see above, the strength of preferences solely matters for the maximum distance between the total group investment and e^*.

Line network: When players are payoff maximizers, every profile is an equilibrium where $(e_{pi} = e^*, e_{ci} = 0, e_{cj} + e_{pj} = e^*)$ for $i \in \{1, 2\}$. Now, consider an ORE and suppose that players' social preferences are compatible and small and that players only compare with their direct neighbors. That is, suppose that $\omega \in \Omega^{line}$ with

$$\begin{aligned} R_\tau &= \mathcal{N}_k(g) \quad \text{for all } \tau \in \mathcal{T}_k, \ k \in \mathcal{N}, \\ \mathcal{T}_i &= \mathcal{T}_i^{**} \quad \text{for all } i \in \mathcal{C}, \\ \mathcal{T}_j &= \mathcal{T}_j^{**} \quad \text{for all } j \in \mathcal{P}, \\ \epsilon &< e^*/5, \end{aligned}$$

and where $\mathcal{T}_i^{**}$ and $\mathcal{T}_j^{**}$ are defined in (9). Then, all refined ORE entail a periphery-sponsored public good for which $\pi(c_i, \omega) \geq \pi(p_i, \omega)$ for all $\omega \in \Omega^{line}$ and $i \in \{1, 2\}$.

Thus, preference compatibility also selects among the ORE in the line network. Yet, it does so less effectively than in the star, core periphery, and d-box because Lemma 2 can only be applied to the line end players and their direct neighbors in the middle of the line.

Circle: The absence of any *nested* neighborhoods in the circle network puts an end to the equilibrium selection property of social preferences. All that can be said about the set of ORE is thus summarized in condition (5): The ORE set is wider than the payoff-maximizing equilibrium set and collapses with it if $\epsilon \to 0$. In particular, when ϵ is small, other-regarding players coordinate on either a (near) *distributed* investment profile, reminiscent of the egalitarian profiles in the complete network, or a (near) *specialized* profile where every second player free rides on the investments of her neighbors.

Maybe surprisingly, a specialized profile can even be supported in the circle network when all players are social-welfare concerned or inequity-averse. The intuition extends immediately from what we said about the role of nestedness in Section 3.3. Even though the contributing players might feel exploited, they maintain their investments for the sake of their own payoffs. The free riders, therefore, receive a total contribution beyond their personal desired level of the public good and consequently see no reason to bear the extra cost of a more equal outcome.

Network ranking: So far, our theory predicted marked differences between the networks in our experiment in terms of how well they facilitate coordination in a group of heterogeneous, socially concerned players. Our theory does predict more, however. It allows for an exact ranking of these networks in terms of how likely they facilitate coordination on a refined ORE set when player groups are randomly assembled.

A first observation is that for a random draw on the entire state set Ω, the likelihood that we yield a combination of compatible preference types declines, ceteris paribus, with the *size* of a network. That type of coordination problem is well known in the literature (e.g., Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000), and it has immediate consequences for the dyad and complete networks:[19]

$$P\big(\omega \in \Omega^{dyad}\big) \;\geq\; P\big(\omega \in \Omega^{comp}\big).$$

There is a second type of coordination problem, however, that is related

[19] In the theory of Fehr and Schmidt (1999), for example, the number of players adversely affects the likelihood of cooperation in public goods games with punishment options because this reduces the likelihood that a sufficient number of conditional cooperators is present.

to the existence and the organization of *nested* neighborhoods in a network. We have seen above that in the absence of any nested neighborhoods, such as in the circle, a network is prone to produce multiple equilibria even when all players have homogeneous preferences. Nevertheless, even between the nested networks in Figure 1, our theory predicts some differences with regard to how likely coordination is expected to succeed. The reason is that there is asymmetry with regard to the ideal number of central players, n_c, who nest other players neighborhoods and the ideal number of peripheral players, n_p, whose neighborhoods are nested. The larger n_c, the more likely it is that there is one central player who is of a social-welfare or altruistic type so that this player would be willing to sponsor the local public good when no one else in her neighborhood does. The reverse is true with n_p. The more peripheral players in a network, the more likely it is that at least one of them is willing to help out a center player who earns less than her. In other words, the likelihood of assembling a group of players with compatible social preferences decreases with the number of central players but increases with the number of peripheral players in a network. This logic leads to the following ranking:[20]

$$\begin{aligned} P\big(\omega \in \Omega^{star_2}\big) &\geq y \in \left\{ P\big(\omega \in \Omega^{core_2}\big)\ ,\ P\big(\omega \in \Omega^{dbox_2}\big) \right\} \\ &\geq P\big(\omega \in \Omega^{line}\big). \end{aligned} \tag{10}$$

Furthermore, coordination is easier in an asymmetric than in a symmetric network:

$$y \in \left\{ P(\omega \in \Omega^{core_2})\,,\, P(\omega \in \Omega^{dbox_2}) \right\} \quad \geq \quad P(\omega \in \Omega^{comp}).$$

Both claims follow immediately from inspection of the conditions leading to equilibrium selection in the different networks. Comparing the star and the d-box, for instance, makes clear that any preference constellation that does *not* meet the compatibility requirements in the star also does not meet the requirements in the d-box. For example, an altruist in the star center position is sufficient to support a center-sponsored public good in an ORE. But one altruist in the center of the d-box has the same effect. Conversely, two competitive players in the peripheral positions of the star are still sufficient for preference compatibility (when the third periphery player is, e.g., inequity-averse), while two competitive players are insufficient in the d-box. Altogether, our theory thus leads to the following testable predictions:

Hypothesis 1: In the networks of Figure 1, except the circle network, a group of players with compatible social preferences is more likely to coordinate on a refined ORE than a group without compatible preferences.

[20] It also holds that $P(\omega \in \Omega^{star_1}) \geq P(\omega \in \Omega^{core_1}) \geq P(\omega \in \Omega^{dbox_1})$ because it is $\overline{\epsilon}^{star} = \overline{\epsilon}^{core} > \overline{\epsilon}^{dbox}$.

Hypothesis 2: The asymmetric networks (star, core periphery, d-box, line), as well as the dyad and complete networks, can be ranked according to how likely a random group of players is going to coordinate on a refined ORE.

Finally, for the circle network, we would expect that even when all the compatibility criteria of Lemma 2 are satisfied by a group of players, the group does nevertheless *not* coordinate more likely on either a specialized or a distributed ORE profile than a group without the proper preference combination.

4.3 Experimental procedure

We administered the experiment in the facilities of the Experimental Laboratory for Sociology and Economics (ELSE) at Utrecht University in The Netherlands. The experiment was programmed in z-tree 3.0 (Fischbacher, 2007) and subjects were recruited via ORSEE (Greiner, 2015).

A total of 120 students participated in eight sessions, with 12–20 students each. No subject could attend more than one session. The average subject's age was 22, 67% were female, and 72% were of Dutch nationality. All subjects played our dynamic extension of the Bramoullé and Kranton (2007) game on each of the seven networks in Figure 1. In each game, subjects were assigned to a random group of players and a random network position. Moreover, the order of the networks was randomly varied between sessions.

Each subject played five games on each network, one trial game and four payoff-relevant games. This means that every subject played 35 games, of which 28 were payoff-relevant. It also means that the experiment comprises a total of 960 payoff-relevant games: 120 games on each of the four-player networks and 240 on the dyad. A typical session lasted 80 minutes, and subjects earned 11.82 euros on average (including a 3 euro show-up fee).

4.4 Social preference elicitation

Key to our testing of Hypotheses 1 is that we have an estimate for the social preference parameters (ρ_s, σ_s) of our subjects. We estimated these parameters directly from their behavior in the network games.[21] Concretely, we assumed that a subject chose an investment e_i at time t of a game so as to maximize the following augmented utility function (4):

$$U_s(e_i, e_{-i}, i) + \theta_{(e_i, t_s, i)}. \tag{11}$$

[21]There are some practical reasons for this. The experiment was already 80 minutes long and additional preference elicitation (dictator) games would have meant burdening students with more time-consuming tasks. In addition, prior research has shown that social preferences can be context-dependent (Tversky and Simonson, 1993; List, 2007; Stoop, Noussair, and Van Soest, 2012), and we thus suspected that other-regarding behavior might change from the "cold" environment of a dictator game to the "heated", interactive environment of our network games.

Here, e_{-i} denotes the momentary investments of the other players and $\theta_{(e_i,t_s,i)}$ a random utility component that captures the impact of other unobserved factors on the choice of e_i. The dependent variable in our econometric model is thus a binary variable that takes the value of one for the investment level that a subject has actually chosen at time t and a value of zero for all alternative investments (limited to $e_i \in \{0, 1, 2, ..., 15\}$). We then computed the $(\hat{\rho}_s, \hat{\sigma}_s)$-pair that maximizes the full likelihood of the actual investments chosen by a subject, where for our main specification we assumed that subjects only include their direct neighbors in their reference group (i.e., $R_s = \mathcal{N}_i(g)$).

In doing so, we confined the set of network games used for our estimations in two ways. First, we ensured that we did *not* use the same games for our estimations that we used to test our theory on. In particular, every time we tested the conformity of a group's play with our theoretical predictions, we estimated the social preferences of the group's members from their investments in some other network games (with different players). For example, for our tests on the fourth repetition of a network game, we made use of the available information in the first three games on the same network and the games on the other six networks. Second, we additionally restricted the set of games to ensure a balanced set of network positions for each subject. This is because the random assignment of players to network positions implies that some positions, particularly the periphery positions of the star network, are over-represented in a subject's set of games compared to, for example, the center position of the star. Based on theory and intuition, we would expect, however, that subjects perceive each position as a different decision situation, triggering a different social comparison concern.[22] We therefore categorized network positions into three classes:

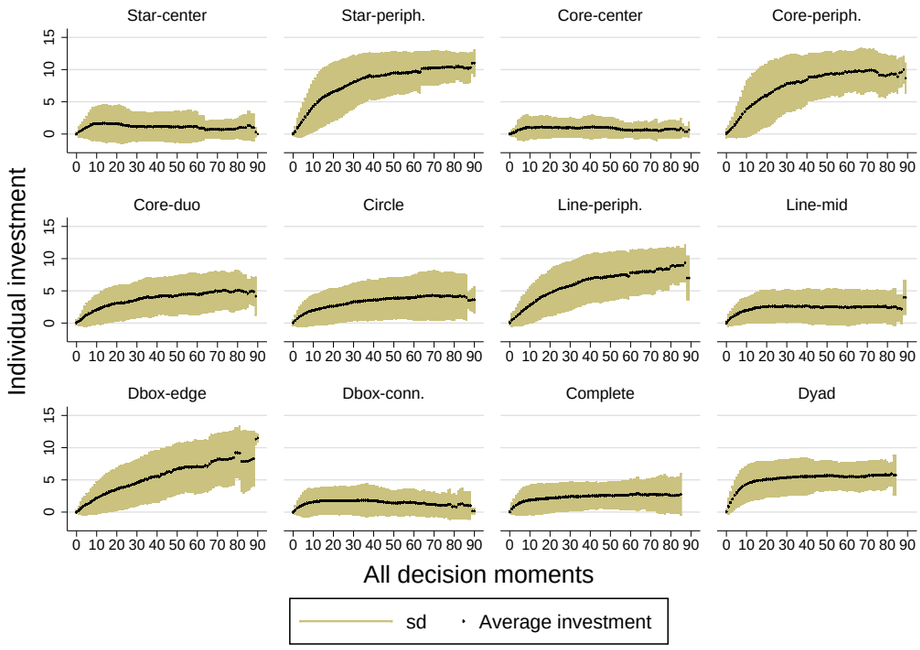
- center positions (of the star, line, core periphery, and d-box)
- periphery positions (of the same networks), and
- symmetric positions (of the circle, core duo, complete network, and dyad).

We then confined our estimations on an equal number of games from each of these classes, whereby we retained the exact order of positions proposed above to ensure that all estimates stem from the same set of positions. Otherwise, we used as many payoff-relevant decision moments ($t \in [30, t^{max}]$) and as many games as possible.[23]

[22] This was confirmed in a pre-test where we estimated the subject-average $(\hat{\rho}_i, \hat{\sigma}_i)$-pair for every network position and found that this average estimate greatly differs by network position.

[23] Obviously, we made several choices. To check the robustness of our findings, we therefore also elicited our subjects' preferences in several alternative ways. Experimental Appendix B.3 summarizes our results on this. Experimental Appendix B.4 then reproduces our main findings based on these alternative estimates.

Figure 3: Investments by network position over time



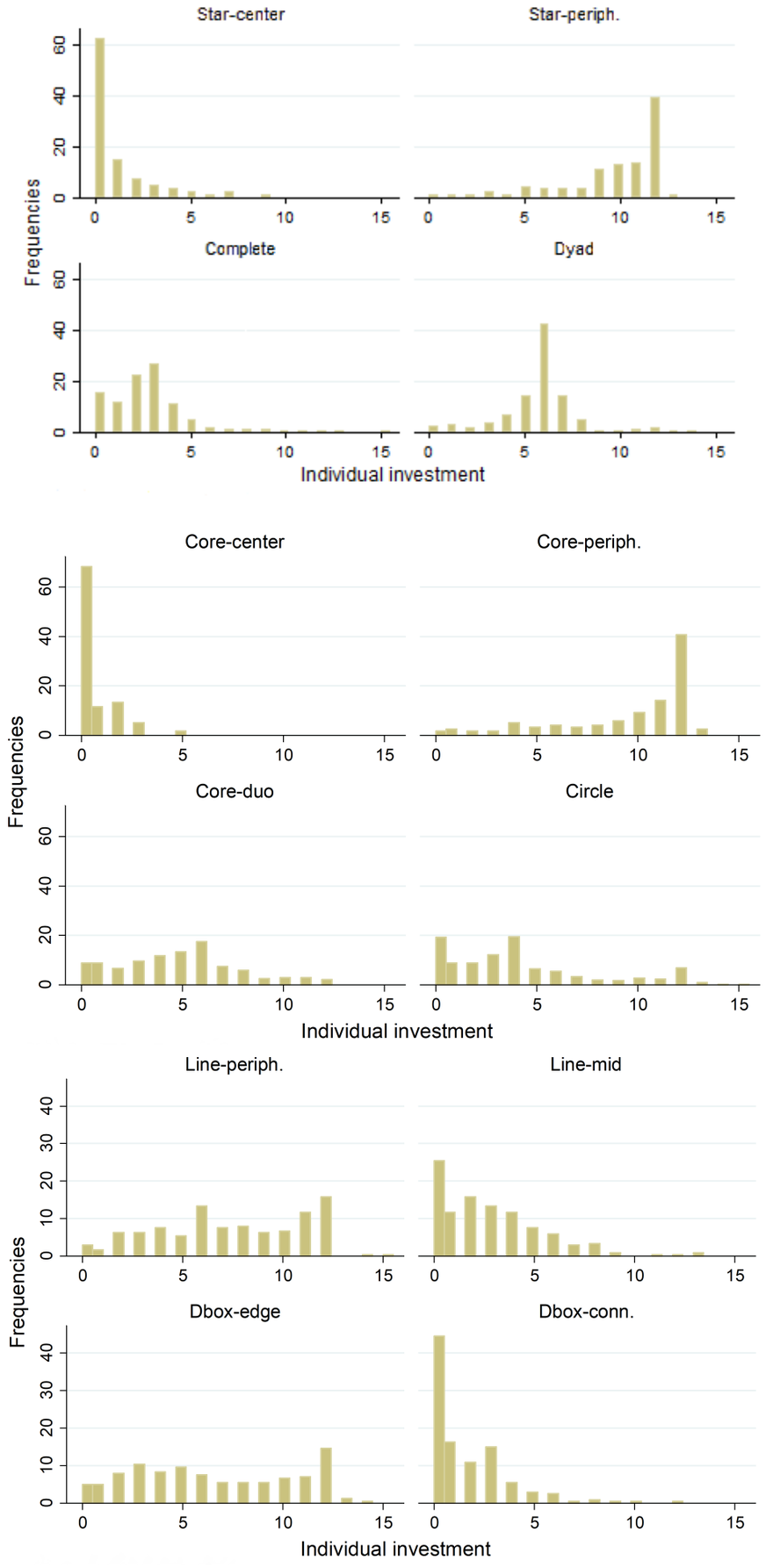
5 Descriptive findings

We first give an overview about the behavior in our experimental games before we test our hypotheses. If not stated otherwise, all results refer to the final investments at the randomly determined game ends.

Position-level findings: Figure 3 plots for each network position the evolution of the average investment and the evolution of the average investment plus/minus one standard deviation over time. Even a glance at the figure suggests that the investments converge to some steady-state value, which is reached roughly between 30 and 70 seconds and which is always lower than the individual payoff-optimum of $e^* = 12$ in all network positions. Thus, the evolution of behavior is reminiscent of some best-response dynamic that converges to a static equilibrium.

In support of this, Figure 4 plots the distributions of the round-end investments per network position. Consistent with our (static) ORE predictions for the dyad and complete networks, the unique distributional modes in these networks are at 6 and 3, respectively, which is consistent with the predicted egalitarian split of $e^* = 12$. Moreover, in line with the predictions for the center positions of the star, core-periphery, d-box, and—to a lesser extent—the line network, the preferred choice is the zero contribution. Subjects in the peripheral positions, in contrast, oftentimes choose $e_j = 12$. Thus, the behavior in all our asymmetric networks is consistent with a periphery-sponsorship equilibrium.

Figure 4: Investments by network position



NOTES: Observations in star center, core center, core periphery, line middle, and line periphery: 120; core duo, d-box center, and d-box edge: 240; star periphery: 360; dyad, complete, and circle: 480. One value in the dyad [29] dropped for better display.

Table 1: Frequencies of other-regarding equilibria

Network	Equilibrium	Deviation from money-maximizing best response: zero ($\epsilon = 0$)	small ($\epsilon < 2$)	mod. ($\epsilon < 3$)	any (any ϵ)
Dyad	egalitarian (rfd)	32.2%	32.2%	46.0%	49.3%
	other	8.8%	25.9%	33.0%	50.7%
Complete	egalitarian (rfd)	0.8%	0.8%	0.8%	0.8%
	other	20.9%	43.4%	62.5%	99.2%
Star	per-spon. (rfd)	15.8%	20.8%	33.3%	62.5%
	cent-sp. with $\pi_c \geq \pi_j$ (rfd)	—	—	—	36.6%
	cent-spon. other	0%	0%	0%	0.8%
Circle	specialized	7.5%	11.6%	16.6%	29.2%
	distributed	3.3%	10.0%	27.5%	70.8%
Core	per-spon. (rfd)	17.5%	34.2%	43.3%	68.3%
	cent-sp. with $\pi_c \geq \pi_j$ (rfd)	—	—	—	31.7%
	cent-spon. other	0%	0%	0%	0%
D-box	per-spon (rfd)	8.3%	11.7%	15.0%	25.8%
	cent-sp. with $\pi_c \geq \pi_j$ (rfd)	—	—	0.8%	64.2%
	cent-spon. other	0%	0%	3.3%	10.0%
Line	end-spon. (rfd)	0.8%	6.7%	15.0%	28.3%
	distr. with $\pi_m \geq \pi_e$ (rfd)	8.3%	12.5%	14.1%	30.8%
	distr. other	1.7%	4.2%	8.4%	40.9%

NOTES: Percentages of investment profiles consistent with an other-regarding equilibrium (ORE) at the random ends of the 960 network games. 240 observations for dyad, 120 for all other networks. Refined ORE are indicated with "(rfd)". The exact criteria for equilibrium consistency are shown in Table 6 in the Theoretical Appendix.

Finally, as expected for the circle network, where our other-regarding theory predicts no selection among the two very different classes of specialized and distributed equilibrium profiles, the distribution of choices has two modes at zero and 12 units (consistent with a specialized equilibrium) and a third mode at four units (consistent with a distributed equilibrium).

Thus, the position-level findings are much in line with our theoretical predictions.[24] Nevertheless, because the investments of all group members need to "fit" in equilibrium, all this is no more than indicative. In the following, we therefore also have a brief look at the group-level behavior.

Group-level findings: Table 1 presents the shares of investment profiles per network that are consistent with a (refined) ORE. The table distinguishes between ORE with four degrees of maximal deviation from a pure payoff-maximizing equilibrium: zero ($\epsilon = 0$), one ($\epsilon < 2$), two ($\epsilon < 3$), and any (any ϵ).[25]

A first observation is that the number of groups converging on a payoff-

[24]Similar pictures emerge when we look at all payoff-relevant decisions on or after the second-30 mark (see Figure 8 in the Experimental Appendix).

[25]The exact consistency requirements for an investment profile are summarized in Table 6 in the Theoretical Appendix. The critical values $\epsilon < 2$ and $\epsilon < 3$ are chosen because a deviation of one (two) units is the maximum deviation for which a periphery-sponsored public good is the unique refined ORE in the d-box and in all the other asymmetric networks, respectively (see Table 6).

maximizing equilibrium (with $\epsilon = 0$) is remarkably high.[26] Not surprisingly, these numbers become even larger when investments in the neighborhood around a payoff-maximizing equilibrium point are rationalized by subjects' social preferences. Interestingly, however, a small expansion of the range of feasible profiles is already enough to capture a significant share of the observed investments. In the periphery positions of star, core periphery, and line network, for example, an inclusion of a deviation of ± 2 units from a payoff-maximizing best response adds meaning to the frequently observed downwards deviations in Figure 4. As a result, the share of profiles consistent with an ORE (with $\epsilon < 3$) more than doubles. Turning to the refined ORE, we first look at Column 3 (with $\epsilon = 0$):

(i) In the asymmetric networks (star, core periphery, d-box, and line), the most frequent ORE profile is a pure periphery-sponsored public good. And, even if some groups coordinate on a partially distributed profile, as in the line, the center players typically earn more. Both are in line with a refined ORE.

(ii) In the dyad network, a large majority of groups (32.2%) splits $e^* = 12$ equally. In contrast, in the complete network, this is the case for only 0.8% of groups (i.e., exactly one group). Thus, our predictions are supported in the dyad but not in the complete network.

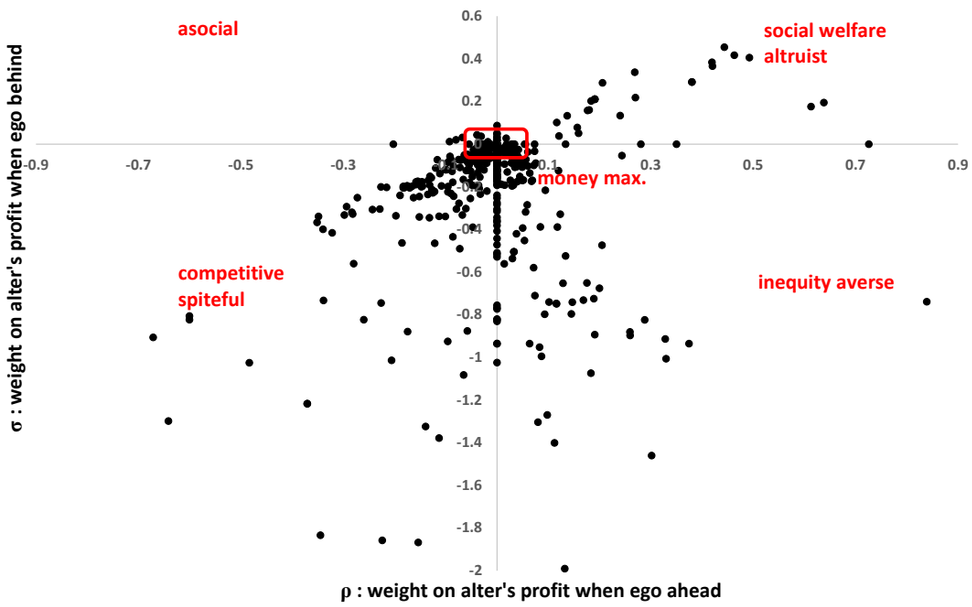
(iii) In the circle network, 7.5% of groups coordinated on a specialized equilibrium with alternating investments of zero and 12 units. Another 3.3% of groups coordinated on an equal-split equilibrium. Thus, as expected, both types of payoff-maximizing equilibria gain experimental support.

Thus, with the exception of the complete network, the numbers in Column 3 are much in line with our refined ORE predictions. Nevertheless, as we will see below, the low share of refined ORE profiles in the complete network makes sense when we take the much stronger preference-compatibility requirements into account.

The same can be said about the wider sets of ORE in Columns 4–6. In the asymmetric networks (star, core, d-box, and line), the vast majority of groups converges to a periphery-sponsored public good. In the dyad, almost half the groups choose an equal-split equilibrium. Finally, in the circle, both the shares of "nearly" specialized profiles and "nearly" distributed profiles increase significantly when we look at the wider sets of ORE. It is only in the complete network where the share of equal-split profiles remains at a low level of 0.8% even when we take all the investment profiles into account where the sum of investment is different from twelve.

[26] Compared to Rosenkranz and Weitzel (2012), for example, who experimentally study the same games but with a non-continuous-time design, the number of groups converging on a payoff-maximizing equilibrium increases by a factor of 3.4 in the star network (smallest increase) to 27 in the circle (largest increase).

Figure 5: Social preference estimates



NOTES: Estimated $(\hat{\sigma}_s, \hat{\rho}_s)$ from subject- and game-specific conditional logit estimations of model (11). See Section 4.4 for procedural details. Ten pairs (with $\hat{\sigma}_s < -2$) are dropped for better display.

To put these findings into perspective, Experimental Appendix B.2 compares the numbers in Table 1 with the predictions of several alternative equilibrium refinement concepts, notably efficiency, Nash tâtonnement stability, and quantal response theory. To sum up the findings, in contrast to efficiency and stability, the power of our social preference theory is that it selects the "natural" equilibria in the dyad and all the asymmetric networks (star, core periphery, d-box, line), that is, an egalitarian equilibrium in the former and a periphery-sponsored public good in the latter. The value-added over quantal response theory is, in turn, that it does not rule out the co-existence of multiple, empirically relevant equilibria.

6 Hypothesis 1: preference compatibility

We have seen above that the experimental data is much in line with our theoretical predictions for most network structures. Nevertheless, the above findings do not rule out the possibility that the data is generated by some other data generating process. Here, we test a discriminatory prediction of our theory that the reason why a subject group coordinates on one of the frequently observed refined ORE is that all its members have a set of compatible social preferences.

Social preference estimates: Towards this end, we first present our social preference estimates for our subject pool and we classify these estimates according to whether they "match" at the group level or not.

Figure 5 summarizes our estimates for (ρ_s, σ_s) from a conditional logit

Table 2: Groups with compatible social preferences

Pref. strength	dyad	star	core	d-box	line	complete
any $\hat{\epsilon}$	27.1%	76.7%	77.5%	62.5%	26.7%	6.7%
$\hat{\epsilon} < 3$	13.3%	20.0%	23.3%	15.8%	8.3%	2.5%
$\hat{\epsilon} < 1$	2.9%	4.2%	3.3%	3.3%	1.7%	0%
No. of groups	240	120	120	120	120	120

NOTES: Categorization of 840 subject groups (all but the 120 groups playing the circle) according to whether their members meet the preference-compatibility requirements of Section 4.2 or not. Groups are additionally classified by the maximum social preference strength of their members.

estimation of our random utility model (11). Our categorization of these estimates into the different social preference types of utility model (4) and the different social preference strength classes of Lemma 1 is summarized in Table 8 in Experimental Appendix B.3.

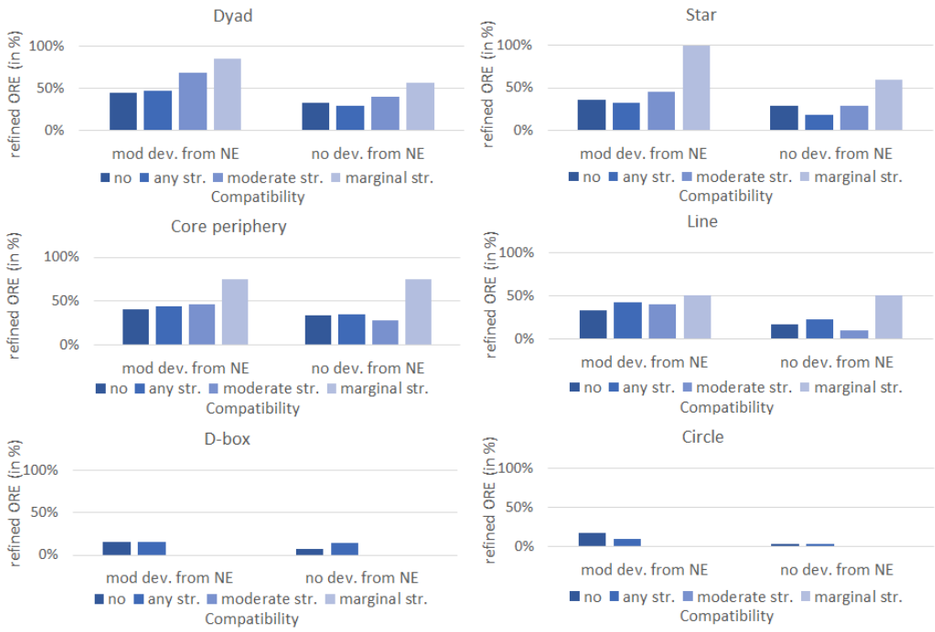
Table 2 presents the results of the next step, the categorization of the subject groups with regard to whether their members have a set of compatible preferences or not. The exact criteria for preference compatibility can be found in Section 4.2. Clearly, for all networks, there is a sizable number of groups that meets the requirements and another sizable number that does not. We can thus turn to our main question: Do groups with compatible social preferences play a refined ORE more often than groups without the proper preference combination?

Hypothesis test: Figure 6 provides some descriptive evidence on this. It shows for each network (except the complete network) the shares of refined ORE in the total number of investment profiles played by groups with compatible and incompatible social preferences, respectively. In particular, we limit our attention to the refined ORE equilibria where investments deviate by no more than ± 2 units from a pure money-maximizing best response simply because many groups coordinated on a refined ORE with any deviation in most of our networks (see Table 1). Moreover, the complete network is omitted from the figure simply because only a single group managed to coordinate on a refined ORE in this network. Panel A looks at all end-game investment profiles. Panel B focuses on the game ends on or after the second-50 mark where investments have reached a steady state in most networks (see Figure 3).

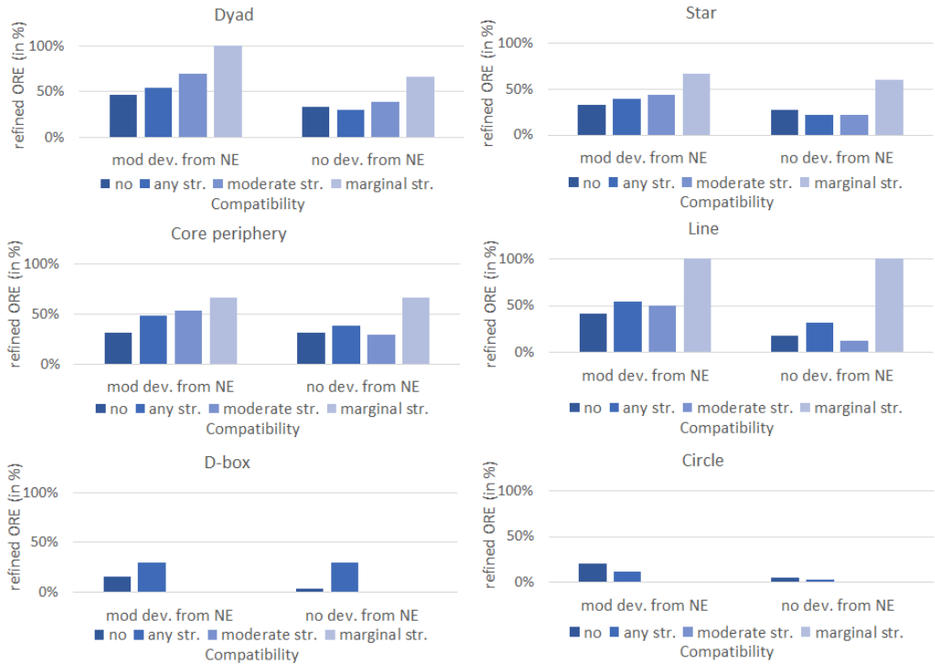
With a few exceptions, the findings are much in line with Hypothesis 1. In the dyad, for example, groups with compatible social preferences clearly coordinated on a refined ORE more often, irrespective of whether we look at all games or only those that ended on or after the second-50 mark. In the star and the core periphery network, we come to the same clear conclusion when we focus on those subject groups where all members have at most a moderate social preference strength. Similarly, in the line network,

Figure 6: Preference compatibility and refined other-regarding equilibria

(a) Panel A



(b) Panel B ($t^{max} \geq 50$)



NOTES: Shares of refined ORE in the total number of investment profiles played by groups with compatible and incompatible social preferences, respectively. The table further distinguishes between compatible groups of different maximal social preference strengths: any social preference strength (any $\hat{e}$), moderate strength ($\hat{e} < 3$), or marginal strength ($\hat{e} < 1$). Refined ORE are measured at two degrees of deviation from a pure money-maximizing equilibrium: moderate deviation ($\epsilon < 3$) and no deviation ($\epsilon = 0$). Panel A: 120 observations per four-player network and 240 for the dyad, from all random game ends. Panel B: 55–88 observation per four-player network and 162 for the dyad, from the random game ends on or after the second-50 mark.

preference compatibility is predictive when we look at groups with a small preference strength or games that ended on or after the second-50 mark. In

contrast, in the d-box, we find at most weak support for our theory. Only when we look at the games that ended late, the expected relationship between preference compatibility and equilibrium selection becomes visible.

These observations can be corroborated in an econometric analysis. In Table 3, we report the results of six multinomial logit models. The dependent variable in all six models is the same *outcome indicator* as in Figure 6 that classifies the observed round-end investment profiles into the ones that are compatible with a refined ORE and the others that are not. Both classes are further partitioned into a total of six sets of profiles that differ by the extent to which the investments deviate from a pure money maximizing best response. Together, these six classes capture all feasible investment profiles in our experiment. The class of non-refined ORE with a large deviation from a payoff-maximizing equilibrium is chosen as the base outcome. The independent variables are the various versions of preference compatibility already used for Figure 6. Next to these, the models include outcome-specific constants and arrays of group- and network-specific control variables to account for other (unobserved) factors that may explain why a certain investment profile is played more often than the base outcome. In other words, our models account among others for the fact that a certain profile is played more often because it is more efficient, strategically more stable, or the basin of attraction of some other unobserved dynamic process.

The results in Table 3 largely confirm Hypothesis 1: groups with compatible social preferences play the refined ORE profiles more often than groups without the proper preference combination. The effect is particularly pronounced for groups with a moderate or marginal preference strength (Models 2 and 3) and for games that ended on or after the second-50 mark (Models 5 and 6). Moreover, as demonstrated in Experimental Appendix B.4, these results are robust with regard to alternative approaches to elicit the social preferences of our subjects and with regard to separate analysis for groups of subjects with at most moderate or marginal social preference strengths.

7 Hypothesis 2: network size and nestedness

We have seen above that preference compatibility facilitates coordination on a refined set of equilibria in our experimental network games. However, how easily equilibrium selection succeeds also depends, according to our theory, on the size and the nestedness of a network structure.

Hypothesis tests: We provide two pieces of evidence supporting the role of network size and nestedness. The first piece comes from the circle network. Remember that, according to our theory, social preferences should not help to coordinate on either a distributed or a specialized equilibrium profile because no neighborhood is nested in the circle network. To put this

Table 3: Test of Hypothesis 1—Multinomial logit results

Compatible pref. of strength...	refined other-regarding eq. ($\epsilon = 0$)	($0 < \epsilon < 3$)	($3 \leq \epsilon$)	non-refined other-regarding eq. ($\epsilon = 0$)	($0 < \epsilon < 3$)	($3 \leq \epsilon$)
Model 1:						
any $\hat{\epsilon}$	0.774**	0.795*	0.934***	-0.022	-0.130	base
	(0.338)	(0.429)	(0.317)	(0.245)	(0.377)	outcome
Model 2:						
$\hat{\epsilon} < 3$	1.003***	1.540***	1.094**	0.429	0.206	–
	(0.315)	(0.378)	(0.448)	(0.419)	(0.872)	–
Model 3:						
$\hat{\epsilon} < 1$	1.814*	1.530	0.615	-12.5***	-12.1***	–
	(1.047)	(1.353)	(0.574)	(1.073)	(1.045)	–
$t^{max} \geq 50$:						
Model 4:						
any $\hat{\epsilon}$	0.932**	0.839*	0.701*	0.150	0.534	–
	(0.472)	(0.437)	(0.381)	(0.416)	(0.697)	–
Model 5:						
$\hat{\epsilon} < 3$	0.578***	1.322***	0.654***	0.226	0.264	–
	(0.206)	(0.345)	(0.187)	(0.645)	(0.864)	–
Model 6:						
$\hat{\epsilon} < 1$	15.64***	15.29***	14.03***	1.367***	1.293***	–
	(0.839)	(0.752)	(0.818)	(0.448)	(0.297)	–

NOTES: Results of six multinomial logit estimations. Models 1–4: 840 observations from final decision moments ($t = t^{max}$) in all network games, but the games on the circle. Models 5–8: 517 observations from final decision moments on or after the second-50 mark. All models include two group-specific *experience* measures (one measuring the position of a game in a session, the other measuring the x-th repetition of the same network game) and measures of network *size* and *clustering*. Standard errors clustered at the session level in parentheses: $^{***}p < 0.01$,** $p < 0.05$,* $p < 0.1$.

to a test, we constructed measures of preference compatibility for the subject groups playing the circle that "worked" in other networks. Concretely, we checked whether a group matches the compatibility requirements for the complete network or the compatibility requirements for the star network. We then checked whether groups with such a plausible preference combination play a distributed or a specialized profile in the circle network more often than groups without such a combination. Our findings on this Placebo test are illustrated in the cross-table for the circle network in Figure 6. The results of a multinomial logit estimation similar to the models in the previous section are presented in Table 12 in Experimental Appendix B.5. In line with our expectations, neither the cross-table nor the regression table suggest a systematic relationship between preference compatibility and the selection of either a distributed or a specialized profile.

The second piece of evidence stems from a comparison of behavior across all networks in our experiment. As argued in the theory section, the net-

Table 4: Frequency of refined other-regarding equilibria

	dyad	star	core	d-box	line	complete	circle
any ϵ	49.3%	99.2%	100.0%	90.0%	59.1%	0.8%	0%
$\epsilon < 3$	46.0%	33.3%	43.3%	15.8%	29.1%	0.8%	0%
$\epsilon = 0$	32.2%	15.8%	17.5%	8.3%	9.1%	0.8%	0%

NOTES: This is an excerpt of the larger Table 1.

works differ markedly in terms of how easily the requirements for a group with compatible social preferences can be met. Accordingly, we would also expect the networks to differ in terms of many groups succeed to coordinate on a refined ORE. Table 4 reproduces the shares of refined ORE per network that we have already seen in Table 1. Consistent with our expectations, the shares are much higher for the two-player dyad network than for the four-player complete network, which supports the expected detrimental impact of network *size*. Also consistent with our theory, the shares of refined ORE are highest in the star and core periphery, intermediate in the d-box and line, and lowest in the complete network. Thus, Table 1 supports the expected conducive impact of *nested* neighborhoods that are concentrated around a single player or a small number of players in a network who nest all other players' neighborhoods. All these observations gain further support from the regression results shown in Table 5. There, the shares of refined ORE are regressed on a number of network statistics that have been found to facilitate coordination in prior experiments, notably network *size* and *clustering* (Berninghaus, Ehrhart, and Keser, 2002; Cassar, 2007; Charness, Feri, Meléndez-Jiménez, and Sutter, 2014). Next to these, the models include a simple measure of network *nestedness*. The latter is a rank variable that sorts the seven networks in our experiment according to the presence and the structure of their nested neighborhoods: (1) circle, (2) dyad/complete, (3) line, (4) d-box/core, (5) star. Strikingly, even our simple *nestedness* measure is highly significant across all specifications with the expected positive effect ($p < .002$).[27] Notably also, unlike in prior experiments, clustering has if at all a negative impact on successful coordination.

Quantitative fit: So far, we produced several pieces of evidence in support of the key mechanisms behind our theory. Here, we briefly investigate its predictive power. In particular, we answer the following question: Suppose we would only have information about the overall preference type distribution of our subject pool available before the start of the experiment. Based on our theory, we could thus calculate for each network how many groups are expected to have a set of compatible preferences, given that sub-

[27] A more informed measure, which takes into account the preference type distribution of the underlying population of players (such as the measure developed for our quantitative predictions below), performs even better. Moreover, the results in Table 5 are robust with regard to the exclusion of the dyad or the circle network from the sample.

Table 5: Test of Hypothesis 2—OLS results

	Share of refined other-regarding equilibria per session					
	(any ϵ)		($\epsilon < 3$)		($\epsilon = 0$)	
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Nestedness		1.172***		0.379***		0.278***
		(0.009)		(0.045)		(0.059)
Size	-0.185*	-0.722***	-0.714***	-0.888***	-0.455**	-0.582***
	(0.086)	(0.086)	(0.158)	(0.151)	(0.164)	(0.159)
Clustering	-0.353***	-0.243***	-0.196***	-0.160***	-0.096**	-0.070*
	(0.017)	(0.017)	(0.031)	(0.032)	(0.031)	(0.032)
Constant	0.931***	0.796***	1.004***	0.961***	0.638***	0.606***
	(0.090)	(0.091)	(0.151)	(0.154)	(0.160)	(0.163)
Observations	56	56	56	56	56	56
R-squared	0.104	0.853	0.216	0.482	0.141	0.377

NOTES: Results of six OLS estimations with 56 observations each: one observation per network (7 networks) per session (8 sessions). *Nestedness* is a rank variable sorting the networks according to their degree and structure of nested neighborhoods (circle=0, dyad/complete=0.25, line=0.5, d-box/core=0.75, star=1), *size* measures the number of players, *clustering* is the network clustering coefficient. Standard errors clustered at the session level in parentheses: $^{***}p < 0.01$,** $p < 0.05$,* $p < 0.1$.

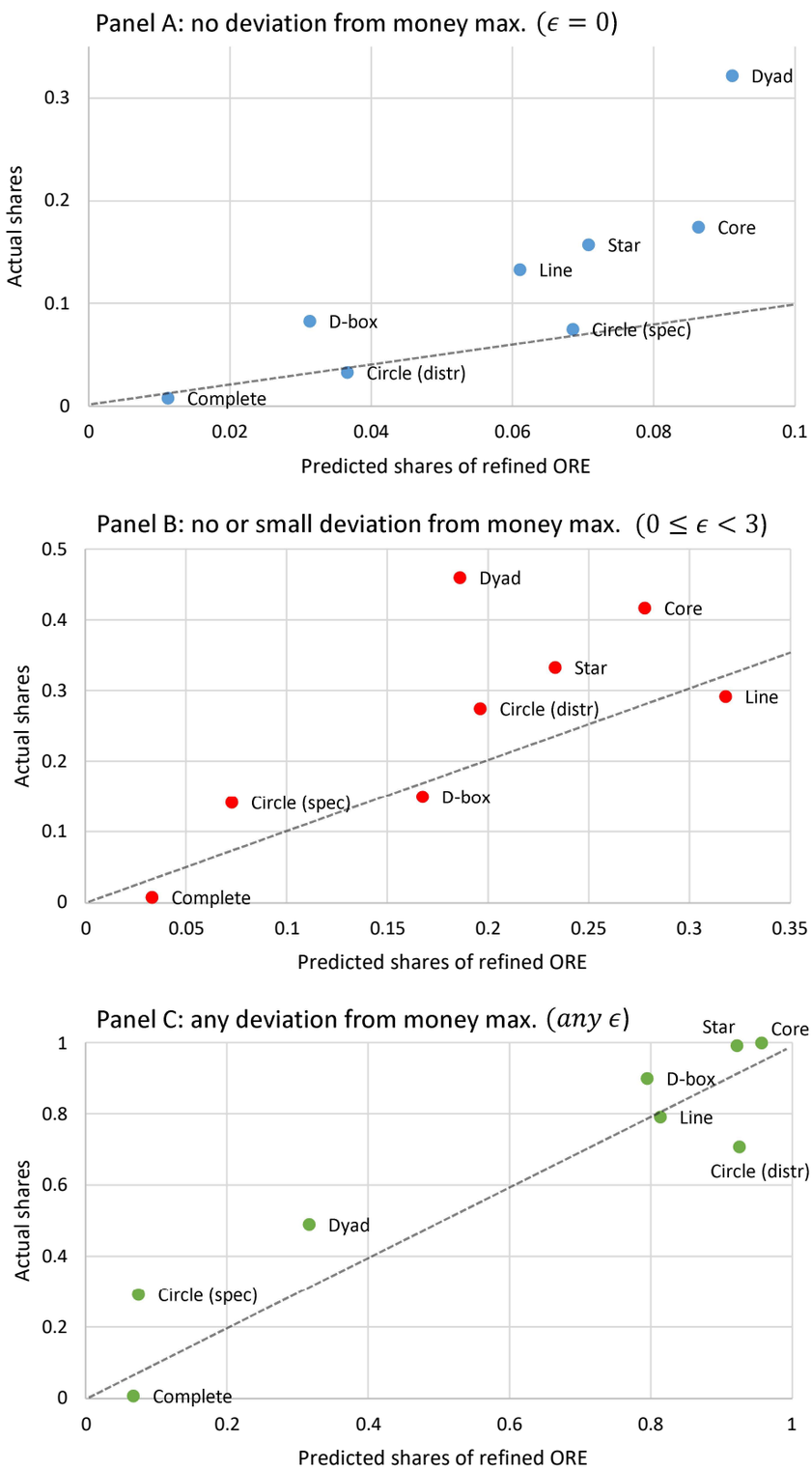
jects are randomly assigned to groups. Could we also make some quantitatively sound prediction for each network about the shares of refined ORE observed at the end of the experiment?

For this purpose, we first have a look at Table 4 on the shares of refined ORE and Table 2 on the number of groups with compatible social preferences. Remember that both tables measure completely different things. The one categorizes the investment profile chosen by a subject group at the end of a game. The other classifies the group members' social preferences, which we elicited from their behavior in other network games. Nevertheless, a careful comparison of the numbers suggests a striking relationship between the number of groups with compatible preferences per network and the shares of refined ORE played on these networks. This begs the question of whether we might be able to predict the shares of refined ORE based on the numbers in Table 2.

Figure 7 summarizes our predictions, which are solely based on Table 2 and three assumptions that immediately from our theory: (i) all groups (compatible or not) choose with certainty an investment profile that corresponds to their social preference strength (that is, groups with marginal preferences ($\hat{\epsilon} < 1$) play a payoff-maximizing equilibrium ($\epsilon = 0$), groups with moderate preference strengths ($1 \leq \hat{\epsilon} < 3$) play a profile with $0 \leq \epsilon < 3$, etc.); (ii) all groups with compatible preferences choose a refined ORE with certainty; (iii) all groups randomize among the remaining investment profiles with equal probability when multiple profiles remain in the subsets generated by (i) and (ii).[28]

[28]The required (conditional) probabilities are summarized in Table 13 of the Experimen-

Figure 7: Actual and predicted shares of other-regarding equilibria



NOTES: Predicted versus actual shares of refined ORE per network for three degrees of deviation from a pure money-maximizing equilibrium. For the circle network, the actual and predicted shares of distributed and specialized ORE are reported. The black dashed line is the 45-degree line.

Overall, the predictions work remarkably well. Even though our model tends to under-predict the frequency of refined ORE (most points lie above

tal Appendix.

the 45-degree line), all three panels indicate a robust positive relationship between the predicted and the actual shares of refined ORE. In particular, the predictions work best for the four-player networks, and the model is capable of explaining the sizable gap between the high shares of refined ORE in the asymmetric networks on one hand and the low share in the complete network on the other.

8 Conclusion

We set out to study how social preferences shape behavior in a network of interdependent social interactions. Do they help to overcome the inequality imposed by a network structure? Do they help to coordinate behaviour when selfish incentives give rise to multiple equilibria? To answer these questions, we developed a model of a local interaction game between socially concerned players. The unique feature of our model is that it incorporates a flexible utilitity function that captures several realistic social preference types.

One of our main results is that social preferences can indeed facilitate equilibrium selection. The key condition for this is that players have the "right", or what we call compatible, preference combination in the sense that players' social preferences fit the network positions they are in. In particular, our theory predicts that coordination is facilitated when the more central network positions are occupied by competitive players and the more peripheral positions by altruists or players with maximin preferences. When this condition is met, then our second important result follows that social preferences do not necessarily produce more equitable or efficient investment profiles in a network. Rather, they reinforce the (in-)equality that is already inherent in a network structure, most notably the symmetry of a fully connected network and the asymmetry of a star-like structure.

Nevertheless, social preferences do not always reinforce the (in-)equality pre-imposed by a network structure. This becomes most obvious from our results on the circle network. Here, our theory predicts that both distributed and specialized equilibria can co-existed. More generally, our third result highlights the importance of nested neighborhoods and the size of the group of peripheral players in these neighborhoods as two crucial properties to facilitate coordination in a network.

Finally, we confirm the predictions of our model in an experiment. Our most interesting finding here is that subject groups with the right preference combination, indeed, played the predicted equilibrium profiles more often than groups without the proper preference combination. Subjects' preferences for a particular network game were thereby estimated from their decisions in all other network games. Another interesting finding is that the variation in nestedness between the networks in our experiment alone is capable of explaining much of the sizable quantitative variation in the shares

of groups that managed to coordinate on the predicted equilibria.

Nevertheless, our sharpest predictions and all our experimental findings concern smaller networks of size four. An open question from this study, therefore, is a more comprehensive investigation of the role of social preferences in larger network games. We were able to show how our equilibria with socially concerned players refine the the money-maximizing equilibria characterized in Bramoullé and Kranton (2007). Yet, a comprehensive comparison of our equilibria with the socially efficient solution to the game or the effect of additional links to a network are beyond what we have achieved. We, therefore, need to leave it for future studies to advance and test our predictions for larger networks.

References

ALLOUCH, N. (2015): "On the private provision of public goods on networks," *Journal of Economic Theory*, 157, 527–552. 2

ANDREONI, J., AND B. D. BERNHEIM (2009): "Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects," *Econometrica*, 77(5), 1607–1636. 2

BELHAJ, M., S. BERVOETS, AND F. DEROÏAN (2016): "Efficient networks in games with local complementarities," *Theoretical Economics*, 11(1), 357–380. 3

BELLEMARE, C., S. KRÖGER, AND A. VAN SOEST (2008): "Measuring inequity aversion in a heterogeneous population using experimental decisions and subjective probabilities," *Econometrica*, 76(4), 815–839. 1, 1, 35

BERNINGHAUS, S. K., K.-M. EHRHART, AND C. KESER (2002): "Conventions and local interaction structures: experimental evidence," *Games and Economic Behavior*, 39(2), 177–205. 2, 7

BERNINGHAUS, S. K., K.-M. EHRHART, AND M. OTT (2006): "A network experiment in continuous time: The influence of link costs," *Experimental Economics*, 9(3), 237–251. 1, 4.1

BOLTON, G. E., AND A. OCKENFELS (2000): "ERC: A theory of equity, reciprocity, and competition," *American Economic Review*, pp. 166–193. 4.2

BONCINELLI, L., AND P. PIN (2012): "Stochastic stability in best shot network games," *Games and Economic Behavior*, 75(2), 538–554. 2

BOURLÈS, R., Y. BRAMOULLÉ, AND E. PEREZ-RICHET (2017): "Altruism in networks," *Econometrica*, 85(2), 675–689. 2, 10

BRAMOULLÉ, Y., A. GALEOTTI, AND B. ROGERS (2016): *The Oxford handbook of the economics of networks*. Oxford University Press. 4

BRAMOULLÉ, Y., AND R. KRANTON (2007): "Public goods in networks," *Journal of Economic Theory*, 135(1), 478–494. 1, 2, 3.1, 3.1, 9, 3.2, 3.3, 12, 3.3, 3.3, 4, 4.1, 4.3, 8, B.2, B.3

BRAMOULLÉ, Y., R. KRANTON, AND M. D'AMOURS (2014): "Strategic interaction and networks," *American Economic Review*, 104(3), 898–930. 2

BRUHIN, A., E. FEHR, AND D. SCHUNK (2019): "The many faces of human sociality: Uncovering the distribution and stability of social preferences," *Journal of the European Economic Association*, 17(4), 1025–1069. 11, B.3, B.3

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. 8

CALLANDER, S., AND C. R. PLOTT (2005): "Principles of network development and evolution: An experimental study," *Journal of Public Economics*, 89(8), 1469–1495. 1, 4.1

CASSAR, A. (2007): "Coordination and cooperation in local, random and small world networks: Experimental evidence," *Games and Economic Behavior*, 58(2), 209–230. 2, 7

CHARNESS, G., F. FERI, M. A. MELÉNDEZ-JIMÉNEZ, AND M. SUTTER (2014): "Experimental games on networks: Underpinnings of behavior and equilibrium selection," *Econometrica*, 82(5), 1615–1670. 2, 7, B.2, 32, 33

CHARNESS, G., AND M. RABIN (2002): "Understanding social preferences with simple tests," *The Quarterly Journal of Economics*, 117(3), 817–869. 1, 3.2, 12, B.3, B.3

CHOI, S., S. KARIV, AND E. GALLO (2016): "Networks in the Laboratory," in *The Oxford Handbook of the Economics of Networks*. 4

COOK, K. S., AND R. M. EMERSON (1978): "Power, Equity and Commitment in Exchange Networks," *American Sociological Review*, 43(5), 721. 2

ELLISON, G. (1993): "Learning, local interaction, and coordination," *Econometrica*, 61(5), 1047–1071. 2

FALK, A., A. BECKER, T. DOHMEN, B. ENKE, D. HUFFMAN, AND U. SUNDE (2018): "Global evidence on economic preferences," *The Quarterly Journal of Economics*, 133(4), 1645–1692. 1, 1

FALK, A., AND M. KOSFELD (2003): *It's all about connections: Evidence on network formation*. Centre for Economic Policy Research. C.1

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. 8

FEHR, E., AND K. M. SCHMIDT (1999): "A theory of fairness, competition, and cooperation," *The Quarterly Journal of Economics*, 114(3), 817–868. 4.2, 19

FISCHBACHER, U. (2007): "z-Tree: Zurich toolbox for ready-made economic experiments," *Experimental Economics*, 10(2), 171–178. 4.3, C.1

GALEOTTI, A., AND S. GOYAL (2010): "The law of the few," *American Economic Review*, 100(4), 1468–92. 5

GALEOTTI, A., S. GOYAL, M. O. JACKSON, F. VEGA-REDONDO, AND L. YARIV (2010): "Network games," *The Review of Economic Studies*, 77(1), 218–244. 2

GHIGLINO, C., AND S. GOYAL (2010): "Keeping up with the neighbors: social interaction in a market economy," *Journal of the European Economic Association*, 8(1), 90–119. 2, 6, 10

GOYAL, S., AND M. C. JANSSEN (1997): "Non-exclusive conventions and social coordination," *Journal of Economic Theory*, 77(1), 34–57. 2

GOYAL, S., S. ROSENKRANZ, U. WEITZEL, AND V. BUSKENS (2017): "Information acquisition and exchange in social networks," *The Economic Journal*, 127(606), 2302–2331. 1, 5

GREINER, B. (2004): "An online recruitment system for economic experiments," . C.1

——— (2015): "Subject pool recruitment procedures: organizing experiments with ORSEE," *Journal of the Economic Science Association*, 1(1), 114–125. 4.3

HARSANYI, J. C., AND R. SELTEN (1988): "A general theory of equilibrium selection in games," *MIT Press Books*, 1. 32

IMMORLICA, N., R. KRANTON, M. MANEA, AND G. STODDARD (2017): "Social status in networks," *American Economic Journal: Microeconomics*, 9(1), 1–30. 2, 10

JACKSON, M. O., AND A. WOLINSKY (1996): "A strategic model of social and economic networks," *Journal of Economic Theory*, 71(1), 44–74. 2

KERSCHBAMER, R., AND D. MÜLLER (2020): "Social preferences and political attitudes: An online experiment on a large heterogeneous sample," *Journal of Public Economics*, 182, 104076. 1, 35

KÖNIG, M. D., C. J. TESSONE, AND Y. ZENOU (2014): "Nestedness in networks: A theoretical model and some applications," *Theoretical Economics*, 9(3), 695–752. 3

LI, X. (2019): "Designing Weighted and Directed Networks Under Complementarities," *Available at SSRN 3299331*. 3

LIST, J. A. (2007): "On the interpretation of giving in dictator games," *Journal of Political Economy*, 115(3), 482–493. 21

MARIANI, M. S., Z.-M. REN, J. BASCOMPTE, AND C. J. TESSONE (2019): "Nestedness in complex networks: observation, emergence, and implications," *Physics Reports*, 813, 1–90. 3

MAURICE, J., A. ROUAIX, AND M. WILLINGER (2013): "Income Redistribution and Public Good Provision: An Experiment," *International Economic Review*, 54(3), 957–975. 8

MCKELVEY, R. D., AND T. R. PALFREY (1995): "Quantal response equilibria for normal form games," *Games and Economic Behavior*, 10(1), 6–38. B.2

OLAIZOLA, N., AND F. VALENCIANO (2020): "Characterization of efficient networks in a generalized connections model with endogenous link strength," *SERIEs*, pp. 1–27. 3

RICHEFORT, L. (2018): "Warm-glow giving in networks with multiple public goods," *International Journal of Game Theory*, 47(4), 1211–1238. 2, 6

RIEDL, A., AND A. ULE (2002): "Exclusion and cooperation in social network experiments," *Unpublished Paper, CREED, University*. 5

ROSENKRANZ, S., AND U. WEITZEL (2012): "Network structure and strategic investments: An experimental analysis," *Games and Economic Behavior*, 75(2), 898–920. 2, 4.1, 17, 26, B.2, B.2

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. 8

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. 8

SCHULZ, U., AND T. MAY (1989): "The recoding of social orientations with ranking and pair comparison procedures," *European Journal of Social Psychology*, 19(1), 41–59. 3.2

SOBEL, J. (2005): "Interdependent Preferences and Reciprocity," *Journal of Economic Literature*, 43(2), 392–436. 1, 4, 3.2

STOOP, J., C. N. NOUSSAIR, AND D. VAN SOEST (2012): "From the lab to the field: Cooperation among fishermen," *Journal of Political Economy*, 120(6), 1027–1056. 21

TVERSKY, A., AND I. SIMONSON (1993): "Context-dependent preferences," *Management science*, 39(10), 1179–1189. 21

ZHANG, Y. (2018): "Social Preferences in Network Games: Theory and Laboratory Evidence," *Available at SSRN 3059710*. 2

Appendices

A Theoretical Appendix

A.1 Missing proofs of general statements

Lemma 1 repeated. Suppose that players' utilities are defined by payoff function (1) and the social preference model (4) with parameters $t_s = (\rho_s, \sigma_s, R_s)$. A player's social preference strength, ϵ_s, is given by

$$\begin{aligned}&\text{for an altruist or a social-welfare type } (\rho_s \geq \sigma_s \geq 0): \ \epsilon_s^p\\ &\text{for an inequity-averse type } (\rho_s > 0 > \sigma_s > -1): \ \max\{\epsilon_s^p\,;\,\epsilon_s^n\}\\ &\text{for a competitive or spiteful type } (0 \geq \rho_s \geq \sigma_s > -1) \ : \ \epsilon_s^n\end{aligned}$$

where

$$\begin{aligned}\epsilon_s^p &= (b')^{-1}\big(\frac{c}{1+\rho_s}\big) - e^* \\ \epsilon_s^n &= e^* - \max\big\{(b')^{-1}\big(\frac{c}{1+\sigma_s}\big)\,;\,0\big\}.\end{aligned} \tag{12}$$

Proof of Lemma 1. For given 'expected' investments in the other nodes,

$$e_{-i} = \big(e_{\tau_1 1}, ..., e_{\tau_t 1}, ..., e_{\tau_1 i-1}, ..., e_{\tau_t i-1}, e_{\tau_1 i+1}, ..., e_{\tau_t i+1}, ..., e_{\tau_1 n}, ..., e_{\tau_t n}\big) \in \mathbb{R}_+^{|\Omega_{-i}|},$$

the first-order condition for an optimal investment of a type t_s in node i is given by[29]

$$\begin{aligned}\frac{\partial U}{\partial e_{t_s i}}(t_s, i, e_{t_s i}, e_{-i}) \ &= \ \sum_{\omega\in\Omega} p(\omega)\Big[b'\big(e_{t_s i} + \sum_{j\in\mathcal{N}_i(g)} e_{\tau j}\big) - c \\ &+ \frac{\sigma_s}{|R_s|}\sum_{j\in R_s^-(\omega)} b'\big(e_{t_s i} + \sum_{k\in\mathcal{N}_j(g)\setminus\{i\}} e_{\tau k}\big) \ + \ \frac{\rho_s}{|R_s|}\sum_{j\in R_s^+(\omega)} b'\big(e_{t_s i} + \sum_{k\in\mathcal{N}_j(g)\setminus\{i\}} e_{\tau k}\big)\Big] \ \leq \ 0\,,\end{aligned} \tag{13}$$

where $R_s^+(\omega)$ ($R_s^-(\omega)$) denotes the ω-dependent set of players $j \in R_s$ with $\pi(\omega, i) > (<)\, \pi(\omega, j)$.

[29] When $\partial U/\partial e_{t_s i}$ is undefined at a point $e_{t_s i}$, the first-order condition can be redefined as $U_i'(e_{t_s i} - x) > 0$ and $U_i'(e_{t_s i} + x) < 0$ for some $x \to 0^+$. The desired point $e_{t_s i}$ then satisfies this alternative condition.

We aim to determine an upper bound for $|f^{or}(t_s, i, e_{-i}) - f^{non}((0,0), i, e_{-i})|$ and define the scalar ϵ_s for this purpose:

$$\epsilon_s \equiv \max\left\{ \left| f^{or}(t_s, i, e_{-i}) - f^{non}((0,0), i, e_{-i}) \right| \forall\, e_{-i} \in \mathbb{R}_+^{|\Omega_{-i}|},\ \forall i \in \mathcal{N},\ \forall g \in G \right\}.$$

By inspection of (13), the absolute deviation from $f^{non}((0,0), i, e_{-i})$ is maximal when player i is the player who earns the least or most in her entire reference group. That is, either set $R_s^+(\omega) = R_s$ or $R_s^-(\omega) = R_s$ for all ω. Moreover, because $b(\cdot)$ is concave, set $e_{\tau j} = 0$ for all $j \neq i$. The first-order condition then becomes

$$0 \;\geq\; \frac{\partial U}{\partial e_{t_s i}}(t_s, e_{t_s}, e_{-i} = 0) \;=\; \begin{cases} b'(e_{t_s})[1+\rho_s] - c & \text{if} \quad R_s^+(\omega) = R_s \\ b'(e_{t_s})[1+\sigma_s] - c & \text{if} \quad R_s^-(\omega) = R_s \end{cases}.$$

The corresponding first-order condition of a payoff maximizer is $b'(e_i) - c = 0$. This gives $f^{non}((0,0), e_{-i} = 0) = e^*$ and

$$f^{or}(t_s, i, e_{-i}) \;=\; \begin{cases} (b')^{-1}\left(\frac{c}{1+\rho_s}\right) & \text{if} \quad R_s^+(\omega) = R_s \\ \max\left\{(b')^{-1}\left(\frac{c}{1+\sigma_s}\right);\, 0\right\} & \text{if} \quad R_s^-(\omega) = R_s \end{cases}.$$

Hence, we get

$$\epsilon_s \;=\; \max\left\{ (b')^{-1}\left(\frac{c}{1+\rho_s}\right) - e^*;\; e^* - \max\left\{(b')^{-1}\left(\frac{c}{1+\sigma_s}\right);\, 0\right\} \right\},$$

which corresponds to the boundaries summarized in Lemma 1. ■

Proposition 2 repeated. Consider a network g with a fully interconnected, but otherwise isolated, component $\mathcal{C}(g')$, with $g' \subseteq g$ such that $\forall i, j \in \mathcal{C}(g')$ and $k \in \mathcal{N}\backslash\mathcal{C}(g')$, $g_{ij} = 1$ and $g_{ik} = 0$. Suppose that all players $i \in \mathcal{C}(g')$ only compare with their direct neighbors (i.e., $R_i = \mathcal{C}(g')\backslash\{i\}$) and that $\epsilon \to 0$. In a refined ORE, it must be

$$e_i = e_j = \frac{e^*}{|C(g')|} \quad \text{for all } i, j \in \mathcal{C}(g').$$

Proof of Proposition 2. Consider either the perfect-information game of the experiment or the incomplete-information game of Section 3.3. Consider moreover the general case of arbitrary social preference strengths, that is, any $\epsilon > 0$.

We prove that in all these cases $e_{\tau_s i} = e_{\tau_t j} \geq 0$ holds for all $\omega \in \Omega^c$, assuming however that all preferences are compatible in Ω^c. That is, we assume that $\Omega^c \equiv \mathcal{T}_i^c \times ... \times \mathcal{T}_n^c$ contains (i) no player of the type {altruist, spite}, (ii) no two or more $\mathcal{T}_i^c$ contain a money maximizer, and (iii) no two or more $\mathcal{T}_i^c$ contain a distinct type from the set {money maximizer, social welfare, competitive}.

Suppose, to the contrary, that $e_{t_s i} > e_{t_t j}$ for some $t_s \in \mathcal{T}_i^c$ and $t_t \in \mathcal{T}_j^c$. Let $e_{t_s i}$ be the highest of all investments and $e_{t_t j}$ be the lowest of all investments. It then follows that $\pi(t_s, i) \leq \pi(\tau, k)$ for all $\tau \in \mathcal{T}_k^c$, $k \neq i$. Likewise, it holds that $\pi(t_t, j) \geq \pi(\tau, k)$ for all $\tau_{\in}\mathcal{T}_k^c$, $k \neq j$.

Moreover, because $e_{t_s i}$ and $e_{t_t j}$ are best-response investments, they must necessarily satisfy

$$\frac{\partial U}{\partial e_{t_s i}}(t_s, i, e_{t_s i}) = \sum_{\omega \in \Omega^c} p(\omega) b'\left(e_{t_s i} + \sum_{k \neq i} e_{\tau k}\right)\left(1 + \frac{|R_s^-(\omega)|}{n-1}\sigma_s\right) - c \;=\; 0\,, \tag{14}$$

$$\frac{\partial U}{\partial e_{t_t j}}(t_t, j, e_{t_t j}) = \sum_{\omega \in \Omega^c} p(\omega) b'\left(e_{t_t j} + \sum_{k \neq j} e_{\tau k}\right)\left(1 + \frac{|R_t^+(\omega)|}{n-1}\rho_t\right) - c \;\leq\; 0\,, \tag{15}$$

where $p(\omega)$ denotes the (conditional) probability of the preference combination

ω, which becomes a degenerate probability in case of a perfect information game. Moreover, $|R_t^+(\omega)|$ denotes the number of types in player j's reference group that have a lower payoff than j in state ω, and $|R_s^-(\omega)|$ the number of types in i's reference group that have a higher payoff than i.

Because $e_{t_s i}$ is the largest investment, it follows that $e_{t_s i} \geq e_{\tau i}$ for all $\tau \in \mathcal{T}_i^c$. Similarly, because $e_{t_t j}$ is the smallest investment, it follows that $e_{t_t j} \leq e_{\tau j}$ for all $\tau \in \mathcal{T}_j^c$. Therefore, $b'(\cdot)$ in (14) is strictly smaller than $b'(\cdot)$ in (15). Furthermore, because at least one of the inequalities $\sigma_s < 0$ or $\rho_t > 0$ hold true, it follows that for the $e_{\tau_s i}$ that solves $\frac{\partial U}{\partial e_{\tau_s i}}(t_s, i, e_{\tau_s i}) = 0$ it holds that $\frac{\partial U}{\partial e_{\tau_t j}}(t_t, j, e_{\tau_s i}) > 0$. This, however, suggests that $e_{\tau_t j} > e_{\tau_s i}$, a contradiction. In a refined ORE, it must therefore hold that $e_{\tau_s i} = e_{\tau_t j}$ for all players and all their (compatible) types.

Concerning the level of $e_{\tau_s i} = e_{\tau_t j} = e$, the best-response condition (5) implies that every single investment should not be 'too far away' from e^*/n. Suppose, for example, that $\mathcal{T}_i^c = \{\text{social welfare, inequity-averse}\}$ for all i. Then, an optimal e must satisfy

$$\lim_{x \to 0^+} \frac{\partial U}{\partial e_{t_s i}}(t_s, i, e + x) = b'\big(e + x + (n-1)e\big)\big(1 + \sigma_s\big) - c \; < \; 0\,,$$
$$\lim_{x \to 0^+} \frac{\partial U}{\partial e_{t_s i}}(t_s, i, e - x) = b'\big(e - x + (n-1)e\big)\big(1 + \rho_s\big) - c \; > \; 0\,,$$

for some $x \to 0^+$ (because $\partial U_i / \partial e_i(\cdot)$ is undefined at point e). Because it is $\sigma_s = 0$ for a social-welfare type, this means that only some $e \geq e^*/n$ can be supported in a refined ORE. Moreover, by condition (5), the upper bound for e is given by $e \leq \frac{e^* + \epsilon}{n}$. ■

A.2 Measuring social preference strength in the experiment

Our experimental games differ from the more general games of Section 3 in two respects: first, we implement a linear-quadratic payoff function with a 'kink' after which the benefit function $b(\cdot)$ becomes linear. Second, our experimental games resemble a game of perfect information about the investments of every other player. These modifications lead to the following refinement of Lemma 1.

Lemma 1 refined. Consider one of the seven networks in Figure 1. Suppose the utility of player i is defined by payoff function (2) with $a = 29$, $c = 5$, and $x = 1$, and by the social preference model (4) with parameters $t_s = (\rho_s, \sigma_s, R_s)$. Suppose furthermore that player i observes the investments e_{-i} of the other players. It then follows that the player's strength of social preference, ϵ_s, is given by

$$\begin{aligned}
&\text{for an altruist or a social-welfare type } (\rho_s \geq \sigma_s \geq 0): \;\; \epsilon_s^p \\
&\text{for an inequity-averse type } (\rho_s > 0 > \sigma_s > -3): \;\; \max\{\epsilon_s^p\,;\, \epsilon_s^n\} \\
&\text{for a competitive or spiteful type } (0 \geq \rho_s \geq \sigma_s > -3) \;: \;\; \epsilon_s^n
\end{aligned}$$

where

$$\begin{aligned}
\epsilon_s^p &= \max\left\{\rho_s \frac{63}{6 + 2\rho_s}\,;\, \frac{63}{2} - \frac{6}{\rho_s}\right\} \\
\epsilon_s^n &= \left|\sigma_s \frac{94.5 - 31.5\sigma_s}{9 + \sigma_s^2}\right|
\end{aligned}$$

Proof. For given 'observed' investments in the other nodes,

$$e_{-i} = \big(e_1, ..., e_{i-1}, e_{i+1}, ..., e_n\big) \in \mathbb{R}_+^{n-1}$$

the first-order condition of an optimal investment of a type t_s in node i is given by[30]

$$\begin{aligned} 0 \quad \geq \quad \frac{\partial U_i}{\partial e_{t_s i}} \quad &= \quad \begin{cases} 24 - 2e_i^* & \text{if } e_i^* \leq 14 \\ -4 & \text{otherwise} \end{cases} \qquad (16) \\ + \quad \frac{1}{|R_s|}\Big(\sum_{j \in R_s^+} \rho_s + \sum_{j \in R_s^-} \sigma_s \Big) &\begin{cases} 29 - 2\sum_{k \in \mathcal{N}_j(g) \cup \{j\}} e_k & \text{if } \sum_{k \in \mathcal{N}_j(g) \cup \{j\}} e_k \leq 14 \\ 1 & \text{otherwise} \end{cases}, \end{aligned}$$

where e_i^* is defined as $e_i^* \equiv e_{t_s i} + \sum_{j \in \mathcal{N}_i(g)} e_j$ and where R_s^+ (R_s^-) denotes the set of players $j \in R_s$ with $\pi_i > (<) \, \pi_j$. We aim to determine an upper bound for $|f^{or}(t_s, i, e_{-i}) - f^{non}((0,0), i, e_{-i})|$:

$$\epsilon_s \equiv \max \left\{ \left| f^{or}(t_s, i, e_{-i}) - f^{non}((0,0), i, e_{-i}) \right| \forall \, e_{-i} \in \mathbb{R}_+^{n-1}, \, \forall i \in \mathcal{N}, \, \forall g \in G \right\}.$$

Clearly, this upper bound can be equivalently determined by

$$\epsilon_s = \max \left\{ \, |e_i^* - e^*| \, : \, \forall \, e_{-i} \in \mathbb{R}_+^{n-1}, \, \forall i \in \mathcal{N}, \, \forall g \in G \right\},$$

where for the experimental payoff function: $e^* = 12$. Clearly also, from the second line in (16), ϵ_s can be found in the sub-domain of $\mathbb{R}_+^{n-1} \times \mathcal{N} \times G$ where $\sum_{k \in \mathcal{N}_j(g) \cup \{j\}} e_k \leq 14$ for all $j \in \mathcal{N}_i(g)$. This is because any value for ϵ_s outside this sub-domain can also be reached by setting $\sum_{k \in \mathcal{N}_j(g) \cup \{j\}} e_k = 14$. Therefore, ϵ_s is given by the largest absolute value of

$$e_i^* - 12 \quad = \quad \begin{cases} \frac{\left(\sum_{j \in R_s^+} \rho_s + \sum_{j \in R_s^-} \sigma_s \right) \left(29 - 2\sum_{k \in \mathcal{N}_j(g) \cup \{j\}} e_k \right)}{2|R_s|} & \text{if } e_i^* \leq 14 \\ e_i^* - 14 + \frac{\left(\sum_{j \in R_s^+} \rho_s + \sum_{j \in R_s^-} \sigma_s \right) \left(29 - 2\sum_{k \in \mathcal{N}_j(g) \cup \{j\}} e_k \right)}{2|R_s|} & \text{otherwise} \end{cases}. \qquad (17)$$

Next, we go through the different possible parameter constellations for (ρ_s, σ_s): Suppose first that $\rho_s \geq \sigma_s \geq 0$ (social-welfare or altruistic type). Then, the largest absolute deviation from $e^* = 12$ consists of a positive deviation so that $|e_i^* - 12|$ is maximal for $R_s^+ = R_s$ and $e_k = 0$ for all $k \notin \mathcal{N}_i(g) \cup \{i\}$. That is,

$$\begin{aligned} \epsilon_s \quad &= \quad \begin{cases} \rho_s \frac{29|R_s| - 2e_i^*}{2|R_s|} - \rho_s \frac{|\mathcal{N}_i(g)| - 1}{|R_s|} e_i & \text{if } e_i^* \leq 14 \\ e_i^* - 14 + \rho_s \frac{29|R_s| - 2e_i^*}{2|R_s|} - \rho_s \frac{|\mathcal{N}_i(g)| - 1}{|R_s|} e_i & \text{otherwise} \end{cases} \\ &= \quad \begin{cases} \rho_s \frac{29|R_s| - 24 - 2(|\mathcal{N}_i(g)| - 1)e_i}{2(|R_s| + \rho_s)} & \text{if } e_i^* \leq 14 \\ \frac{29|R_s| - 24 - 2(|\mathcal{N}_i(g)| - 1)e_i}{2} - \frac{2|R_s|}{\rho_s} & \text{otherwise} \end{cases}. \end{aligned}$$

Because $\partial \epsilon_s / \partial e_i < 0$, set e_i to its smallest feasible value of $e_i = 0$ to get that

$$\epsilon_s \quad = \quad \begin{cases} \rho_s \frac{29|R_s| - 24}{2(|R_s| + \rho_s)} & \text{if } \rho_i \leq \frac{4|R_s|}{29|R_s| - 28} \\ \frac{29|R_s| - 24}{2} - \frac{2|R_s|}{\rho_s} & \text{otherwise} \end{cases}.$$

Furthermore, because $\partial \epsilon_s / \partial |R_s| > 0$, set $|R_s| = 3$ to get that

$$\epsilon_s \quad = \quad \begin{cases} \rho_s \frac{63}{6 + 2\rho_s} & \text{if } \rho_s \leq \frac{12}{59} \\ \frac{63}{2} - \frac{6}{\rho_s} & \text{otherwise} \end{cases}, \qquad (18)$$

which is what we have claimed.

Suppose now that $\rho_s > 0 > \sigma_s > -3$ (inequity aversion). Then, $|e_i^* - 12|$ is

[30] When $\partial U / \partial e_{t_s i}$ is undefined at a point $e_{t_s i}$, the first-order condition can be redefined as $U_i'(e_{t_s i} - x) > 0$ and $U_i'(e_{t_s i} + x) < 0$ for some $x \to 0^+$. The desired point $e_{t_s i}$ then satisfies this alternative condition.

maximal for either expression (18) or for expression (17) with $R_s^- = R_s$ and $e_i^* \leq 14$. In the latter case, ϵ_s is given by the largest absolute value of

$$\begin{aligned} e_i^* - 12 &= \sigma_s \frac{29|R_s| - 2\sum_{j\in R_s}\sum_{k\in\mathcal{N}_j(g)\cup\{j\}} e_k}{2|R_s|} \\ &= \sigma_s\big(\frac{29}{2} - \frac{\sum_{k\notin\mathcal{N}_i(g)\cup\{i\}} e_k}{|R_s|} - \frac{1}{|R_s|}e_i^* - \frac{|\mathcal{N}_i(g)|-1}{|R_s|}e_i\big) \\ &= \sigma_s\left(\frac{29|R_s| - 24}{2(|R_s|+\sigma_s)} - \frac{\sum_{k\notin\mathcal{N}_i(g)\cup\{i\}} e_k}{|R_s|+\sigma_s} - \frac{|\mathcal{N}_i(g)|-1}{|R_s|+\sigma_s}e_i\right), \end{aligned}$$

with the additional constraint that e_i and e_k are such that $\pi_i < \pi_j$ for all $j \in R_s$. It therefore must be $e_i > 0$. In fact, $\hat{e}_i$ is, by (16), given by

$$\hat{e}_i = \frac{24 - 2e_i^*}{2(|\mathcal{N}_i(g)|-1)} + \sigma_s \frac{29|R_s| - 2e_i^*}{2|R_s|(|\mathcal{N}_i(g)|-1)}.$$

Combined, this gives

$$e_i^* - 12 = \sigma_s\left(\frac{29|R_s| - 24}{2(|R_s|+\sigma_s)} - \frac{\sum_{k\notin\mathcal{N}_i(g)\cup\{i\}} e_k}{|R_s|+\sigma_s} - \frac{12 - e_i^*}{|R_s|+\sigma_s} - \sigma_s\frac{29|R_s| - 2e_i^*}{2|R_s|(|R_s|+\sigma_s)}\right).$$

Thus, in order to minimize $e_i^* - 12$, set $e_k = 0$ to get that

$$e_i^* - 12 = \sigma_s \frac{\big(14.5|R_s| - 12\big)\big(|R_s| - \sigma_s\big)}{|R_s|^2 + \sigma_s^2}.$$

Further, because it is $\partial(e_i^* - 12)/\partial|R_s| < 0$, set $|R_s| = 3$ to find that

$$e_i^* - 12 = \sigma_s \frac{94.5 - 31.5\sigma_s}{9 + \sigma_s^2}, \tag{19}$$

which is what we have claimed.

Suppose finally that $0 \geq \rho_s > \sigma_s > -3$ (competitive or spiteful type). Then, $|e_i^* - 12|$ is maximal for (19). ■

A.3 Experimental predictions

The following predictions characterize the refined ORE of both the more general game of Section 3 with incomplete information as well as the game of perfect information implemented in the experiment. For both characterizations, we essentially make use of condition (5) on the maximal deviation from a payoff-maximizing best response and Lemma 2 on the ordering of payoffs in a refined ORE. The sole difference is that the set of ORE is somewhat smaller in the perfect information game because players can respond to the actual investments of their neighbors and, therefore, do not need to rely on their expectations about other players' investments, for which we can only specify some upper and lower constraints. For example, while a periphery player in the star network is predicted to invest a value e_{t_sj} from the interval $e^* - 3\epsilon_s \leq e_{t_sj} \leq e^* + 3\epsilon_s$ in the incomplete-information version of our game, this interval is no larger than $e^* - \epsilon_s \leq e_{t_sj} \leq e^* + \epsilon_s$ in the case of perfect information.

Other-regarding equilibria in the star, core periphery, and d-box: We first show that an ORE entails either a (near) center- or a (near) periphery-sponsored public good in the incomplete information case.

Suppose first that $e_{\tau c} = 0$ for all types playing the center position(s) (periphery sponsorship). An (unconstrained) payoff maximizer in the periphery position would respond with $f^{non}\big((0,0),p\big) = e^*$. By condition (5), a social type would

therefore respond with

$$e^* - \epsilon \leq f^{or}\big(\tau, p, e_{-j}\big) \leq e^* + \epsilon\,.$$

For the responses of social types in the duo positions of the core periphery network, we additionally need to understand how they respond to each other. Note, however, that for the linear-quadratic payoff function (2) and for a payoff maximizer in duo position j it holds that $f^{non}\big((0,0), j\big) = e^* - \mathbb{E}[e_{\tau d_{-j}}]$. At the same time, it holds that

$$\mathbb{E}[e_{\tau d_j}] \leq \max\{f^{or}\big(\tau, j, e_{-j}\big)\} = e^* - \mathbb{E}[e_{\tau d_{-j}}] + \epsilon$$

because the expression on the right-hand side defines the maximum investment of any social type in duo position j. Adding $\mathbb{E}[e_{\tau d_{-j}}]$ to both sides, we get that $\mathbb{E}[\sum_{j\in\mathcal{D}} e_{\tau d_j}] \leq e^* + \epsilon$. By a similar argument, we also get that $\mathbb{E}[\sum_{j\in\mathcal{D}} e_{\tau d_j}] \geq e^* - \epsilon$. Together, this implies that the joint investments of two money maximizers in the duo positions of the core periphery network satisfy

$$2e^* - (e^* + \epsilon) \leq \sum_{j\in\mathcal{D}} f^{non}\big((0,0), j, e_{-j}\big) \leq 2e^* - (e^* - \epsilon)\,,$$

and the joint investments of two social types in these positions:

$$2e^* - (e^* + \epsilon) - 2\epsilon \leq \sum_{j\in\mathcal{D}} f^{or}\big(\tau, j, e_{-j}\big) \leq 2e^* - (e^* - \epsilon) + 2\epsilon\,.$$

Next, suppose that $e_c > 0$ for at least one social type in the center position(s) of the star, core periphery, and d-box (center sponsorship). An (unconstrained) payoff maximizer would invest

$$f^{non}\big((0,0), c\big) = e^* - \mathbb{E}[\sum_{j\neq c} e_{\tau j}]\,.$$

Payoff maximizers in the periphery positions of all these networks and payoff maximizers in the duo positions of the core periphery would in turn invest

$$\begin{aligned} f^{non}\big((0,0), p\big) &= e^* - \mathbb{E}[\sum_{i\in\mathcal{C}} e_{\tau i}] \quad \text{and} \\ f^{non}\big((0,0), d_j\big) &= e^* - \mathbb{E}[e_{\tau c}] - \mathbb{E}[e_{\tau d_{-j}}]\,. \end{aligned}$$

Combined with condition (5) regarding the best response of a social type, this means that the expected investments are constrained by

$$\begin{aligned} e^* - \mathbb{E}[\sum_{j\neq c} e_{\tau j}] - \epsilon \quad &\leq \mathbb{E}[e_{\tau c}] \leq \quad e^* - \mathbb{E}[\sum_{j\neq c} e_{\tau j}] + \epsilon\,, \\ e^* - \mathbb{E}[\sum_{i\in\mathcal{C}} e_{\tau i}] - \epsilon \quad &\leq \mathbb{E}[e_{\tau p}] \leq \quad e^* - \mathbb{E}[\sum_{i\in\mathcal{C}} e_{\tau i}] + \epsilon\,, \\ e^* - \mathbb{E}[e_{\tau c}] - \mathbb{E}[e_{\tau d_{-j}}] - \epsilon \quad &\leq \mathbb{E}[e_{\tau d_j}] \leq \quad e^* - \mathbb{E}[e_{\tau c}] - \mathbb{E}[e_{\tau d_{-j}}] + \epsilon\,. \end{aligned}$$

For the star network, this gives

$$e^* - 2\epsilon \leq \mathbb{E}_s[e_{\tau c}] \leq e^* + 2\epsilon \quad \text{and} \quad 0 \leq \mathbb{E}_s[e_{\tau p_j}] \leq 3\epsilon.$$

For the d-box, this gives

$$e^* - 5\epsilon \leq \mathbb{E}_d[\sum_{i\in\mathcal{C}} e_{\tau i}] \leq e^* + \epsilon \quad \text{and} \quad 0 \leq \mathbb{E}_d[\sum_{j\in\mathcal{P}} e_{\tau j}] \leq 4\epsilon.$$

And for the core periphery, this gives

$$e^* - 3\epsilon \leq \mathbb{E}_c[e_{\tau c}] \leq e^* + \epsilon \quad \text{and} \quad 0 \leq \mathbb{E}_c[e_{\tau p}] \leq 4\epsilon \quad \text{and} \quad 0 \leq \mathbb{E}_c[\sum_{j\in\mathcal{D}} e_{\tau j}] \leq 4\epsilon.$$

Continuing from here, we determine the upper and lower boundaries for the in-

vestments of a payoff maximizer and, based on condition (5), for a social type in these positions. For the star, this gives

$$e^* - 10\epsilon \leq f_s^{or}(\tau, c, e_{-i}) \leq e^* + \epsilon \quad \text{and} \quad 0 \leq f_s^{or}(\tau, p_j, e_{-j}) \leq 3\epsilon\,. \tag{20}$$

For the d-box, this gives

$$e^* - 11\epsilon \leq \sum_{i \in \mathcal{C}} f_d^{or}(\tau, i, e_{-i}) \leq e^* + 7\epsilon \quad \text{and} \quad 0 \leq f_d^{or}(\tau, p_j, e_{-j}) \leq 6\epsilon\,. \tag{21}$$

And for the core periphery network, this gives

$$\begin{aligned} e^* - 9\epsilon \leq f_c^{or}(\tau, c, e_{-i}) \leq e^* + \epsilon \quad \text{and} \quad 0 \leq f_c^{or}(\tau, p, e_{-j}) \leq 4\epsilon \\ \text{and} \quad 0 \leq f_c^{or}(\tau, d_j, e_{-j}) \leq 4\epsilon\,. \end{aligned} \tag{22}$$

We next derive the boundaries for an ORE in the perfect information game implemented in the experiment.

Players' beliefs about the expected investment of a player in any node i are given by $\mathbb{E}[e_{\tau i}] = e_i$ in this game. This means that for an ORE with $e_c = 0$ in the center position(s) of the star, core periphery, and d-box it must still hold that $e^* - \epsilon \leq f^{or}(\tau, p, e_{-j}) \leq e^* + \epsilon$ in the periphery positions. In the duo position of the core periphery, we get, in contrast, the modified boundaries

$$e^* - \epsilon \leq \sum_{j \in \mathcal{D}} f^{or}(\tau, j, e_{-j}) \leq e^* + \epsilon\,.$$

For an ORE with $e_c > 0$ for at least one player in the center position(s), we get that

$$e^* - \sum_{j \neq c} e_j - \epsilon \quad \leq e_c \leq \quad e^* - \sum_{j \neq c} e_j + \epsilon\,, \tag{23}$$

$$e^* - \sum_{i \in \mathcal{C}} e_i - \epsilon \quad \leq e_p \leq \quad e^* - \textstyle\sum_{i \in \mathcal{C}} e_i + \epsilon\,, \tag{24}$$

$$e^* - e_c - e_{d_{-j}} - \epsilon \quad \leq e_{d_j} \leq \quad e^* - e_c - e_{d_{-j}} + \epsilon\,. \tag{25}$$

It follows from (23) that $\sum_{i \in \mathcal{N}} e_i \leq e^* + \epsilon$ and from (24) and (25) that $\sum_{i \in \mathcal{C}} e_c + e_p \geq e^* - \epsilon$ and $e_c + \sum_{j \in \mathcal{D}} e_{d_j} \geq e^* - \epsilon$. In combination, we find that the periphery players in the star and the d-box, except for periphery player p, jointly contribute at most

$$\sum_{j \in \mathcal{P} \setminus \{p\}} e_j \quad = \quad \sum_{j \in \mathcal{P}} e_j + \sum_{i \in \mathcal{C}} e_i - \big[\sum_{i \in \mathcal{C}} e_c + e_p\big] < e^* + \epsilon - \big[e^* - \epsilon\big] = 2\epsilon.$$

Drawing the same comparison for any other periphery player p_2, we again find that $\sum_{j \in \mathcal{P} \setminus \{p_2\}} e_j \leq 2\epsilon$. Hence, the total contribution received by the center(s) is at most

$$\sum_{j \in \mathcal{P}} e_j \leq \sum_{j \in \mathcal{P} \setminus \{p\}} e_j + \sum_{j \in \mathcal{P} \setminus \{p_2\}} e_j \leq 4\epsilon\,.$$

Similarly, in the core periphery, we find for the periphery player and the duo players, respectively, that

$$\begin{aligned} e_p \quad &= \quad \sum_{i \neq c} e_i + e_c - \big[e_c + e_p\big] \leq e^* + \epsilon - \big[e^* - \epsilon\big] = 2\epsilon\,, \\ \sum_{j \in \mathcal{D}} e_j \quad &= \quad \sum_{i \neq c} e_i + e_c - \big[e_c + \sum_{j \in \mathcal{D}} e_j\big] \leq e^* + \epsilon - \big[e^* - \epsilon\big] = 2\epsilon. \end{aligned}$$

The the total contribution received by the center player in the core periphery network is thus, again, at most $\sum_{j \in \mathcal{D}} e_j + e_p < 4\epsilon$. For the peripheral player with the lowest contribution in the star, the core periphery, and the d-box, (24) requires that $\min_p\{e_p\} + \sum_{i \in \mathcal{C}} e_i \geq e^* - \epsilon$. Thus, the centers players' contributions are necessar-

ily larger than

$$\sum_{i\in\mathcal{C}} e_i \geq e^* - \epsilon - \min_p\{e_p\} \geq e^* - \epsilon - \frac{4\epsilon}{n-|\mathcal{C}|},$$

whereby the lower bound is determined by a situation where all peripheral (and duo) players equally share 4ϵ. Moreover, (23) implies that the center players' contributions are necessarily smaller than

$$\sum_{i\in\mathcal{C}} e_i \leq e^* + \epsilon\,.$$

Together, this defines the boundaries on the investments in a center-sponsored public good in Table 6.

We finally show that when players' preferences are compatible and small, *no* center-sponsored public good can be maintained in a *refined ORE*.

Note that independent of the information that players have, the investments of the center player(s) $i \in \mathcal{C}$ converge to

$$\lim_{\epsilon\to 0} f^{or}\big(\tau, i, e_{-i}\big) = e^*\,.$$

At the same time, the investments of all other players $j \notin \mathcal{C}$ converge to

$$\lim_{\epsilon\to 0} f^{or}\big(\tau, j, e_{-j}\big) = 0\,.$$

Thus, there exists an $\overline{\epsilon}^h$ in the star, core periphery, and d-box such that for $\epsilon < \overline{\epsilon}^h$ it holds that $\pi(i,\omega) < \pi(j,\omega)$ for all possible type combinations $\omega \in \Omega^h$. A contradiction to Lemma 2 because the required payoff ordering can only be maintained when $\pi(i,\omega) \geq \pi(j,\omega)$ for all $i \in \mathcal{C}$ and $j \in \mathcal{N}\backslash\mathcal{C}$ and at least one $\omega \in \Omega^h$.[31]

The critical value $\overline{\epsilon}^h$ depends on the network structure in the following way: In the incomplete information game, it follows from (20) for the star network that

$$\pi_s(i,\omega) \leq b\big(e^*\big) - c(e^* - 9\epsilon)$$

and that

$$\pi_s(j,\omega) \geq \min\left\{b\big(e^* - 10\epsilon\big)\,;\, b\big(e^* + 4\epsilon\big) - 3c\epsilon\right\}.$$

Hence, $\overline{\epsilon}^{star}$ is defined by

$$\max\{\epsilon\}: \quad c \geq \frac{b(e^*) - \min\left\{b\big(e^* - 10\epsilon\big)\,;\, b\big(e^* + 4\epsilon\big) - 3c\epsilon\right\}}{e^* - 9\epsilon}.$$

From (21), it follows that $\overline{\epsilon}^{core}$ is defined by

$$\max\{\epsilon\}: \quad c \geq \frac{b(e^* + 3\epsilon) - \min\left\{b\big(e^* - 9\epsilon\big)\,;\, b\big(e^* + 5\epsilon\big) - 4c\epsilon\right\}}{e^* - 9\epsilon},$$

and from (22), it follows that $\overline{\epsilon}^{dbox}$ is defined by

$$\max\{\epsilon\}: \quad c \geq 2\frac{b(e^* + \epsilon) - \min\left\{b\big(e^* - 11\epsilon\big)\,;\, b\big(e^* + 13\epsilon\big) - 6c\epsilon\right\}}{e^* - 11\epsilon}.$$

A comparison of the critical values gives $\overline{\epsilon}^{star} > \overline{\epsilon}^{core} > \overline{\epsilon}^{dbox}$.

In the perfect information case, similar considerations lead to $\overline{\epsilon}^{star} = \overline{\epsilon}^{core}$, which are in turn defined by

$$\max\{\epsilon\}: \quad c \geq \frac{b(e^*) - \min\left\{b\big(e^* - \frac{7}{3}\epsilon\big)\,;\, b\big(e^* + \frac{5}{3}\epsilon\big) - 4c\epsilon\right\}}{e^* - 4\epsilon}.$$

[31] This condition is necessary regardless of whether players compare with everyone else, as in Ω^{h_1}, or only with their network neighbors, as in Ω^{h_2}.

Furthermore, $\overline{\epsilon}^{dbox}$ is defined by

$$\max\{\epsilon\}: \quad c \geq \frac{b(e^*) - \min\left\{b\left(e^* - 3\epsilon\right); b\left(e^* + \epsilon\right) - 4c\epsilon\right\}}{e^* - 4\epsilon}.$$

This gives the ranking $\overline{\epsilon}^{star} = \overline{\epsilon}^{core} > \overline{\epsilon}^{dbox}$. ■

Other-regarding equilibria in the dyad and complete networks: See proof of Proposition 2.

Other-regarding equilibria in the line: We first characterize the set of ORE in the incomplete information case.

There are two refinements on the set of ORE, beyond the general requirements of condition (5). First, suppose that every player has small social preferences and suppose that this is common knowledge. Concretely, suppose $\epsilon < e^*/5$. It follows that all ORE fall into either the class of *periphery-sponsored* profiles or the class of *partially distributed* profiles:

$$\begin{aligned}(per.spon): &\quad \left([e^* - 3\epsilon, e^* + \epsilon]\,,\, [0, 4\epsilon]\,,\, [0, 4\epsilon]\,,\, [e^* - 3\epsilon, e^* + \epsilon]\right),\\ (distr): &\quad \left([e^* \pm \epsilon]\,,\, 0\,,\, e_{m2} + e_{e2} \in [e^* \pm 3\epsilon]\right).\end{aligned}$$

To show this, we fix the sequence of players in the order $e1 - m1 - m2 - e2$ and exclude out-of-equilibrium profiles:

a) Obviously, *no* investment profile can be supported in an ORE in which all types of *three or more* players invest nothing.

b) There are three possible ORE constellations where all types of *two* players invest nothing:

$$\begin{aligned}(i): &\quad \left([e^* \pm \epsilon]\,,\, 0\,,\, 0\,,\, [e^* \pm \epsilon]\right),\\ (ii): &\quad \left([e^* \pm \epsilon]\,,\, 0\,,\, [e^* \pm \epsilon]\,,\, 0\right),\\ (iii): &\quad \left(0\,,\, [e^* - \epsilon, \infty)\,,\, [e^* - \epsilon, \infty)\,,\, 0\right).\end{aligned}$$

Profiles (i) and (ii) are contained in the classes of ORE described above. Concerning (iii), the sum of player $m1$'s and $m2$'s investments must, by condition (5) and by the fact that the payoff function is linear-quadratic, be weakly less than $e^* + 3\epsilon$. Hence, this is *not* an ORE when $2(e^* - \epsilon) > e^* + 3\epsilon$ and thus when $\epsilon < e^*/5$.

c) There are two potential ORE configurations where all types of *one* player invest nothing:

$$\begin{aligned}(iv): &\quad \left([e^* \pm \epsilon]\,,\, 0\,,\, e_{\tau m_2} + e_{\tau e_2} \in [e^* \pm 3\epsilon]\right),\\ (v): &\quad \left(0\,,\, [e^* - \epsilon, \infty)\,,\, e_{\tau m_2}\,,\, e_{\tau e_2}\right).\end{aligned}$$

The first one is contained in the classes of ORE described above. The second one is *not* an equilibrium when for player m_2 it holds that

$$\min_{\omega \in \Omega}\Big\{\sum_{i \in \mathcal{N}} e_{\tau i}\Big\} = e^* - \epsilon + e^* - 3\epsilon \quad > \quad \max_{\omega \in \Omega}\left\{e_{\tau m_2}\right\} = e^* + \epsilon$$

and, thus, when $\epsilon < e^*/5$.

d) When all types of *all* players make a positive contribution, it follows from the best-response conditions of the end players that

$$e^* - \epsilon \leq \mathbb{E}[e_{\tau e_i}] + \mathbb{E}[e_{\tau m_i}] \leq e^* + \epsilon\,.$$

At the same time, the best-response conditions of the middle players imply that

$$e^* - \epsilon \leq \mathbb{E}[e_{\tau e_i}] + \mathbb{E}[e_{\tau m_i}] + \mathbb{E}[e_{\tau m_j}] \leq e^* + \epsilon\,.$$

Together, this gives $0 \leq \mathbb{E}[e_{\tau m_i}] \leq 2\epsilon$ and $e^* - 3\epsilon \leq \mathbb{E}[e_{\tau e_i}] \leq e^* + \epsilon$. From here, we get that

$$e^* - 2\epsilon \leq f^{non}((0,0), e_i) \leq e^* \quad \text{and} \quad 0 \leq f^{non}((0,0), m_i) \leq e^* - (e^* - 3\epsilon)\,,$$

and in turn we get that

$$e^* - 3\epsilon \leq f^{or}(\tau, e_i, e_{-i}) \leq e^* + \epsilon \quad \text{and} \quad 0 \leq f^{or}(\tau, m_i, e_{-i}) \leq 4\epsilon\,.$$

Hence, we arrive at a profile that is contained in the classes of ORE described above.

Next, we characterize the set of ORE in the perfect information case. Suppose that $\epsilon < e^*/3$. Then, all ORE fall into one of the following two classes

$$\begin{aligned} (per.spon): &\quad \big([e^* - 3\epsilon, e^* + \epsilon]\,,\, [0, 2\epsilon]\,,\, [0, 2\epsilon]\,,\, [e^* - 3\epsilon, e^* + \epsilon]\big), \\ (distr): &\quad \big([e^* \pm \epsilon]\,,\, 0\,,\, e_{m2} + e_{e2} \in [e^* \pm \epsilon]\big). \end{aligned}$$

To show this, we again exclude out-of-equilibrium profiles:

a) Obviously, *no* investment profile can be supported in which all types of *three or more* players invest nothing.

b) Straightforward extensions of the arguments above show that of the three possible ORE constellations where all types of *two* players invest nothing, only the following two remain as possible ORE when $\epsilon < e^*/3$:

$$\begin{aligned} (i): &\quad \big([e^* \pm \epsilon]\,,\, 0\,,\, 0\,,\, [e^* \pm \epsilon]\big), \\ (ii): &\quad \big([e^* \pm \epsilon]\,,\, 0\,,\, [e^* \pm \epsilon]\,,\, 0\big). \end{aligned}$$

c) There are two potential ORE configurations where all types of *one* player invest nothing:

$$\begin{aligned} (iv): &\quad \big([e^* \pm \epsilon]\,,\, 0\,,\, e_{\tau m_2} + e_{\tau e_2} \in [e^* \pm \epsilon]\big), \\ (v): &\quad \big(0\,,\, [e^* - \epsilon, \infty)\,,\, e_{\tau m_2}\,,\, e_{\tau e_2}\big). \end{aligned}$$

The first one is contained in the classes of ORE described above. The second one is *not* an equilibrium when for player m_2 it holds that

$$\min\big\{\sum_{i \in \mathcal{N}} e_i\big\} = e^* - \epsilon + e^* - \epsilon \quad > \quad \max\big\{e_{m_2}\big\} = e^* + \epsilon$$

and, thus, when $\epsilon < e^*/3$.

d) When all types of *all* players make a positive contribution, it follows from the best-response conditions of the end players that

$$e^* - \epsilon \leq e_{e_i} + e_{m_i} \leq e^* + \epsilon\,.$$

At the same time, the best-response conditions of the middle players imply that

$$e^* - \epsilon \leq e_{e_i} + e_{m_i} + e_{m_j} \leq e^* + \epsilon\,.$$

Together, this gives $0 \leq e_{m_i} \leq 2\epsilon$ and $e^* - 3\epsilon \leq e_{e_i} \leq e^* + \epsilon$. Hence, we arrive at a profile that is contained in the classes of ORE described above.

Finally, we characterize the set of *refined ORE* under the conditions that (i) players only include their direct neighbors in their reference group, i.e., $R_s = \mathcal{N}_i(g)$; (ii) their preferences are compatible; (iii) their social preferences are small. That is, suppose that $\mathcal{T}_i = \{\text{competitive, spiteful}\}$, $i \in \mathcal{M}$, and $\mathcal{T}_j = \{\text{payoff max., social welfare, altruist}\}$, $j \in \mathcal{E}$. Moreover, suppose that $\epsilon < e^*/7$ in the incomplete information game and $\epsilon < e^*/5$ in the perfect information game.

Start with the incomplete information case: It immediately follows from $\epsilon < e^*/7$ that every middle player earns necessarily more than every end player in a periphery-sponsored ORE. From the compatibility of preferences and $R_s = \mathcal{N}_i(g)$, it also follows that $\pi(m_1, \omega) \geq \pi(e_1, \omega)$ and $\pi(m_2, \omega) \geq \pi(e_2, \omega)$ for all $\omega \in \Omega^{line}$ in a distributed ORE. This is because suppose, to the contrary, that $e_{\tau_s m_2} > e_{\tau_t e_2}$ for some $\omega \in \Omega^{line}$. Take the most investing type t_s in the middle of the line network and the least investing type t_t at the end of the line. It follows that t_s earns less than t_t (and less than every type of the other middle player). However, it must also hold that

$$\frac{\partial U}{\partial e_{t_s m_2}}(t_s, m_2, e_{t_s m_2}) = \sum_{\omega \in \Omega^{line}} p(\omega) b'\big(e_{t_s m_2} + e_{\tau e_2}\big)\big(1 + \frac{|R_s^-(\omega)|}{n-1}\sigma_s\big) - c \;=\; 0\,, \quad (26)$$

$$\frac{\partial U}{\partial e_{t_t e 2}}(t_t, e_2, e_{t_t e_2}) = \sum_{\omega \in \Omega^{line}} p(\omega) b'\big(e_{t_t e 2} + e_{\tau m_2}\big)\big(1 + \frac{|R_t^+(\omega)|}{n-1}\rho_t\big) - c \;\leq\; 0\,. \quad (27)$$

Because $e_{t_s m_2} \geq e_{\tau m_2}$ and $e_{t_t e_2} \leq e_{\tau e_2}$ for all ω, it follows that $b'(\cdot)$ in (26) is smaller than $b'(\cdot)$ in (27). Moreover, because either the middle player is behindness averse or the end player is aheadness averse, or both, it also holds $\sigma_s < 0$ and $\rho_t \geq 0$. Together, this means that for the $e_{t_s m_2}$ that solves $\frac{\partial U}{\partial e_{t_s m_2}}(t_s, m_2, e_{t_s m_2}) = 0$ it holds that $\frac{\partial U}{\partial e_{t_t e_2}}(t_t, e_2, e_{t_s m_2}) > 0$. This, however, implies that $e_{t_s m_2} < e_{t_t e_2}$, a contradiction.

Next, we look at the perfect information case: Every middle player earns necessarily more than every end player in a periphery-sponsored ORE when $\epsilon < e^*/5$. Expanding on the arguments made above, it also follows that $\pi(m_1, \omega) \geq \pi(e_1, \omega)$ and $\pi(m_2, \omega) \geq \pi(e_2, \omega)$ in a distributed ORE when preferences are compatible and $R_s = \mathcal{N}_i(g)$. This is because all that changes is that $p(\omega)$ becomes a degenerate probability in (26) and (27).

Thus, in a refined ORE, it must be $\pi(m_1, \omega) \geq \pi(e_1, \omega)$ and $\pi(m_2, \omega) \geq \pi(e_2, \omega)$ for all $\omega \in \Omega^{line}$. ■

Other-regarding equilibria in the circle: Consider either a game with perfect information or with incomplete information. Suppose that all players' social preferences are small ($\epsilon < e^*/5$ for both games) and suppose that this is common knowledge (in the incomplete information game). We show that the only classes of ORE resemble a *specialized* or a *fully distributed* investment profile.

Let us fix the sequence of players in the order $i - j - k - l$ for this purpose. First, suppose that $e_{\tau m} > 0$ for at least some $\tau \in \mathcal{T}_m$ but for all $m \in \mathcal{N}$ (a fully distributed profile). Based on the best-response condition (5), $e_{\tau m} > 0$ must lie inside the boundaries $\underline{e} \leq e_{\tau m} \leq \bar{e}$, where $\underline{e}$ and $\bar{e}$ are defined by

$$\begin{aligned} \underline{e} \;+\; 2\,\bar{e} \;&=\; e^* - \epsilon\,, \\ \bar{e} \;+\; 2\,\underline{e} \;&=\; e^* + \epsilon\,. \end{aligned}$$

Solving this set of identities and simplifying, we arrive at $\frac{e^*}{3} - \epsilon \leq e_{\tau m} \leq \frac{e^*}{3} + \epsilon$.

Next, suppose that $e_{\tau i} = 0$ for all types of player i. It follows that (at least some types of) player i's neighbors, j and l, must make some positive investment. In particular, it must be $e_{\tau j} > 0$ and $e_{\tau l} > 0$ for at least some types because suppose, to the contrary, that $e_{\tau j} = 0$ (or $e_{\tau l} = 0$, or both are equal to zero) for all types. Then, $e_{\tau k} > 0$ for at least some type since otherwise $e_{\tau i} + e_{\tau j} + e_{\tau k} = 0$ for all $\omega \in \Omega$. In fact, we would require that simultaneously it holds $\mathbb{E}[e_{\tau k}] \geq e^* - \epsilon$ and $\mathbb{E}[e_{\tau l}] \geq e^* - \epsilon$. This, however, leads to a contradiction because it implies for player k: $f^{non}((0,0), k) \leq e^* - (e^* - \epsilon)$ and thus $f^{or}(\tau, k, e_{-k}) \leq 2\epsilon$. However,

since we need to have $E[e_{\tau k}] \leq f^{or}(\tau, k, e_{-k})$, this constellation is at odds with the assumption $\epsilon < e^*/3$.

Thus, if $e_{\tau i} = 0$ for all types of player i, it must hold that $e_{\tau j} > 0$ and $e_{\tau l} > 0$ for at least some of the types of players j and l. But this also implies that $e_{\tau k} = 0$ for all $\tau \in \mathcal{T}_k$ because suppose, to the contrary, that $e_{\tau k} > 0$ for some τ. Because

$$e^* - \epsilon \leq \mathbb{E}[e_{\tau j}] + E[e_{\tau k}] + E[e_{\tau l}] \leq e^* + \epsilon$$

and

$$e^* - \epsilon \leq \mathbb{E}[e_{\tau k}] + E[e_{\tau l}] \leq e^* + \epsilon\,,$$

it follows that $E[e_{\tau l}] \leq 2\epsilon$ and similarly that $E[e_{\tau j}] \leq 2\epsilon$. This implies, however, that the expected total contribution received by player i is no larger than 4ϵ. Hence, for $\epsilon < e^*/5$ it is $E[e_{\tau j}] + E[e_{\tau l}] \leq 4\epsilon < e^* - \epsilon$. A contradiction to $e_{\tau i} = 0$. Thus, it must be $e_{\tau k} = 0$ for all $\tau \in \mathcal{T}_k$ (a specialized equilibrium). In particular, together with the equilibrium investments of j and l, we get that $(0\,,\,[e^* \pm \epsilon]\,,\,0\,,\,[e^* \pm \epsilon])$. ■

Table 6: Experimental predictions

	Payoff-maximizing equilibria	Other-regarding equilibria
Dyad or complete	$\sum_{i\in\mathcal{N}} e_i = 12$ (S,E) (Q: $e_i = e_j = \frac{12}{n}$)	$\sum_{i\in\mathcal{N}} e_i \in [12 \pm \epsilon]$ Compatible social preferences: $e_i = e_j \in [\frac{12}{n} \pm \frac{\epsilon}{n}]$
Star	(i) $e_c = 0, e_p = 12$ (ii) $e_c = 12, e_p = 0$ (S,Q: (i) selected) (E: (ii) selected)	(i) $e_c = 0, e_p \in [12 \pm \epsilon]$ (ii) $e_c \in [12 - \frac{7\epsilon}{3}, 12 + \epsilon]$, $\sum_{j\in\mathcal{P}} e_j \leq 4\epsilon$ Compatible social preferences: $\pi_c \geq \min_{j\neq c}\{\pi_j\}$ Small pref. ($\epsilon < 3$) in addition: (i) selected
Core periphery	(i) $e_c = 0, e_p = 12$, $\sum_{j\in\mathcal{D}} e_j = 12$ (ii) $e_c = 12, e_{-c} = 0$ (S: (i) selected) (Q: (i) selected with $e_d = 6$) (E: (ii) selected)	(i) $e_c = 0, e_p \in [12 \pm \epsilon]$, $\sum_{j\in D} e_j \in [12 \pm \epsilon]$ (ii) $e_c \in [12 - \frac{7\epsilon}{3}, 12 + \epsilon]$, $\sum_{j\neq c} e_j \leq 4\epsilon$ Compatible social preferences: $\pi_c \geq \min_{j\neq c}\{\pi_j\}$ Small pref. ($\epsilon < 3$) in addition: (i) selected
D-box	(i) $e_c = 0, e_p = 12$ (E) (ii) $e_p = 0, \sum_{i\in C} e_i = 12$ (E) (S,Q: (i) selected)	(i) $e_c = 0, e_p \in [12 \pm \epsilon]$ (ii) $\sum_{i\in C} e_i \in [12 - 3\epsilon, 12 + \epsilon]$, $\sum_{j\in\mathcal{P}} e_j \leq 4\epsilon$ Compatible social preference : $\pi_c \geq \min_{j\neq c}\{\pi_j\}$ Small pref. ($\epsilon < 2$) in addition: (i) selected
Line	(i) $e_{ei} = 12, e_{mi} = 0$, $e_{mj} + e_{ej} = 12$ (S) (ii) $e_{mj} = 0, e_{ej} = 12$ (Q) (iii) $e_{mj} = 12, e_{ej} = 0$ (E)	$\forall i : e_i + \sum_{j\in N_i(g)} e_j \geq e^* - \epsilon$ Compatible social preferences: $\pi_m \geq \min_{j\in R_m}\{\pi_j\}$ Small pref. ($\epsilon < 3$) in addition: (i) $e_{ei} \in [12 \pm \epsilon], e_{mi} = 0$, $e_{mj} + e_{ej} \in [12 \pm \epsilon]$, and $\pi_{mj} \geq \pi_{ej}$ (ii) $e_e \in [12 - 3\epsilon, 12 + \epsilon]$, $e_m \leq 2\epsilon$
Circle	(i) $e_i = 0, e_{i+1} = 12$ (ii) $e_i = 4$ (S,E: (i) selected) (Q: (ii) selected)	$\forall i : e_i + e_{i-1} + e_{i+1} > e^* - \epsilon$ Small preferences ($\epsilon < 3$): (i) $e_i = 0, e_{i+1} \in [12 \pm \epsilon]$ (ii) $e_i \in [4 \pm \epsilon]$

NOTES: (Other-regarding) equilibrium predictions for the experimental games where players have perfect information about other players' investments and end-game decisions are rewarded according to payoff function (2) with $e^* = 12$. For comparison, the equilibria selected by other equilibrium refinement methods are indicated as well: (S) asymptotic *stability*, (Q) *quantal* response equilibria with marginal decision errors, and (E) *efficient* equilibra.

B Experimental Appendix

B.1 Additional descriptive evidence

Figure 8: Investments by network position from second 30 to end

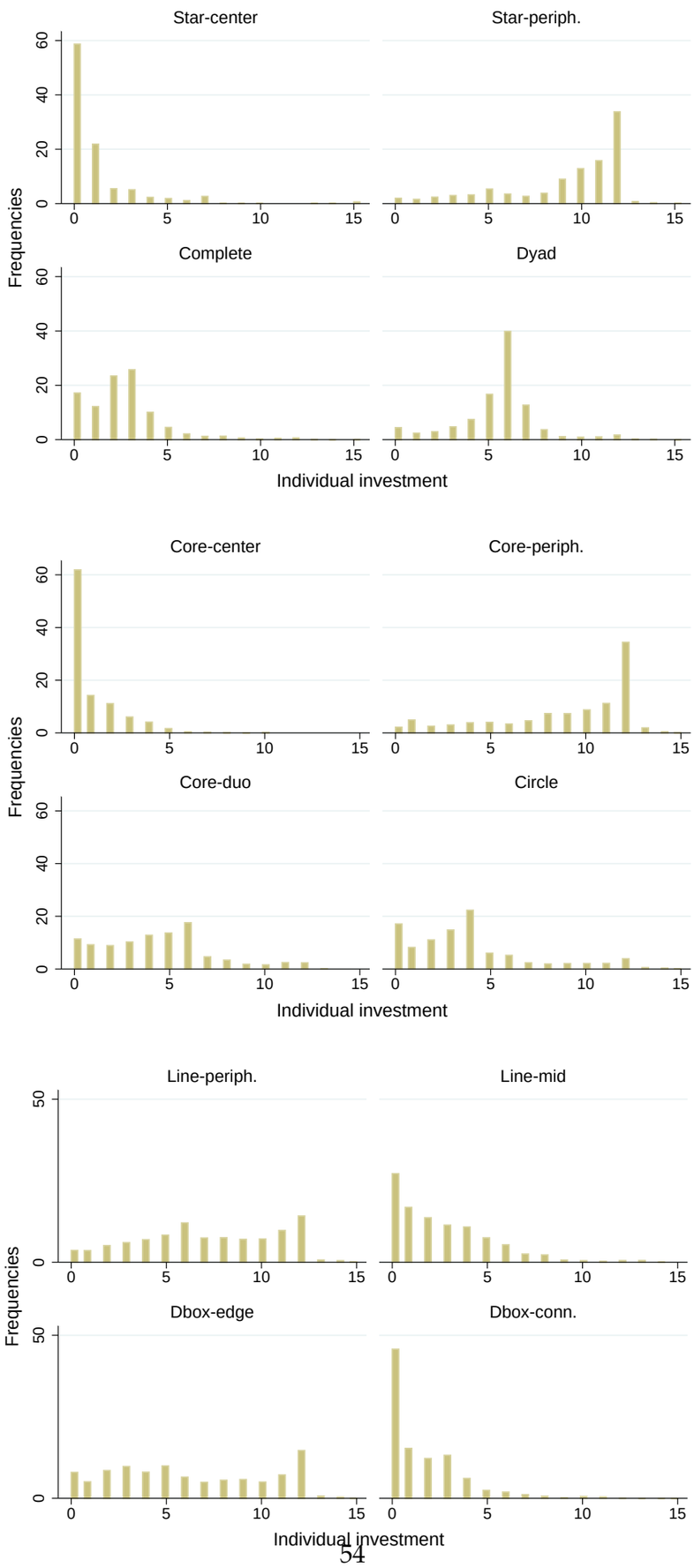


Table 7: Frequency of other-regarding equilibria and refined equilibria

Network	Equilibrium	Deviation from payoff-max. best response: zero ($\epsilon = 0$)	moderate ($\epsilon < 3$)	any (any ϵ)
Dyad	equal	32.2% (S,E,Q,rfd)	46.0% (rfd)	49.3% (rfd)
	other	8.8% (S,E)	33.0%	50.7%
Complete	equal	0.8% (S,E,Q,rfd)	0.8% (rfd)	0.8% (rfd)
	other	20.9% (S,E)	62.5%	99.2%
Star	per-spon	15.8% (S,Q,rfd)	33.3% (rfd)	62.5% (rfd)
	cent-sp: $\pi_c \geq \pi_j$	—	—	36.6% (rfd)
	cent-sp: other	0% (E)	0%	0.8%
Circle	spec	7.5% (S,E)	16.6%	29.2%
	distr	3.3% (Q)	27.5%	70.8%
Core	per-spon	17.5% (S,Q: 4.2%,rfd)	43.3% (rfd)	68.3% (rfd)
	cent-sp: $\pi_c \geq \pi_j$	—	—	31.7% (rfd)
	cent-sp: other	0% (E)	0%	0%
D-box	per-spon	8.3% (S,E,Q,rfd)	15.0% (rfd)	25.8% (rfd)
	cent-sp: $\pi_c \geq \pi_j$	—	0.8% (rfd)	64.2% (rfd)
	cent-sp: other	0% (E)	3.3%	10.0%
Line	end-spon	8.3% (S,Q: 0.8%,rfd)	15.0% (rfd)	28.3% (rfd)
	distr: $\pi_m \geq \pi_e$	8.3% (S,rfd)	14.1% (rfd)	30.8% (rfd)
	distr: other	1.7% (S,E: 0.8%)	8.4%	40.9%

NOTES: Percentages of equilibrium profiles at the random ends of the 960 network games. Observations for dyad: 239 (1 outlier with value 29 dropped). Observations for all other networks: 120. Refined equilibria are: (Q) quantal response, (S) stable, (E) efficient, (rfd) refined other-regarding equilibria. In a few cases, quantal response theory (Q) and efficiency (E) selects a subset of the displayed equilibrium types. The frequencies are reported in parentheses in these cases.

B.2 Comparison with alternative refinement concepts

To put our findings into perspective, this appendix compares our *refined ORE* predictions with those of several alternative equilibrium refinement concepts applied to the Bramoullé and Kranton (2007) game. Table 7 summarizes the predictions of the most relevant concepts:[32]

- *Asymptotic stability* (Bramoullé and Kranton, 2007) based on the idea that stable equilibria might occur more frequently in our continuous-time experiment because a best-response dynamic leads back to them after a single mistake.
- *Efficiency* (Charness, Feri, Meléndez-Jiménez, and Sutter, 2014) based on the idea that subjects might have used the time we gave them to coordinate on a group welfare-maximizing equilibrium.
- *Quantal response* (logit) equilibria (McKelvey and Palfrey, 1995; Rosenkranz and Weitzel, 2012) based on the idea that subjects play a best response to the fluctuating, probabilistic choices of their neighbors.

As can be seen from Table 7, the alternative concepts do not explain our experimental findings better than our social preference theory. On the contrary, they perform worse in some networks:

[32]We omit *risk dominance* as a selection concept (Harsanyi and Selten, 1988; Charness, Feri, Meléndez-Jiménez, and Sutter, 2014) because we deemed it less relevant for our experiment. Subjects were continuously informed about the investments of their group members so that strategic uncertainty is only a minor issue.

- The predictive power of *efficiency* is particularly low in the star and the core periphery network, where it is efficient when the center player provides the public good all by himself. Such a center-sponsored profile is, however, never observed in the data.[33]

- Asymptotic *stability* predicts somewhat better than efficiency, in particular in the star, the core periphery, and the d-box. But it fails to select the empirically relevant equal-split equilibria in the dyad and the end-sponsored public good in the line network simply because all equilibrium profiles are equally stable in these networks.[34]

- Only *quantal response* theory comes close to our social preference theory. As shown in Rosenkranz and Weitzel (2012), the theory selects a unique payoff-maximizing equilibrium given that players make small decision errors. The resulting refined payoff-maximizing equilibria are identical to our refined ORE in all the networks mentioned above. Yet, quantal response theory tends to generate a too fine-grained selection on the equilibrium set. This leads to the situation that in the circle network, for example, quantal response theory predicts an egalitarian split of $e^* = 12$ as the unique equilibrium profile, even though a specialized equilibrium is at least equally relevant in the data.

Thus, in contrast to efficiency and asymptotic stability, the power of our social preference theory is that it selects the "natural" equilibria in the dyad and all the asymmetric networks (star, core periphery, d-box, line), that is, an egalitarian equilibrium in the former and a periphery-sponsored public good in the latter. The value-added over quantal response theory is, in turn, that it does not rule out the co-existence of multiple, empirically relevant equilibria.

B.3 Comparison with other social preference estimates

Table 8 categorizes our preferred social parameter estimates—based on the type classification implied by utility model (4) and Lemma 1—into seven distinct preference types and three distinct strength classes.

The preference *strength* estimates are much in line with our observed deviations from a pure payoff-maximizing best response in Figure 4. In particular, 68.9% of the estimates imply a moderate ($\hat{e}_s < 3$) and 42.4% no more than a marginal ($\hat{e}_s < 1$) social preference strength. This is in line with the observed deviations in Figure 4, confirming the validity of our estimations. Yet, Table 8 reveals some marked differences between our estimated preference *types* and those reported in earlier studies, in particular the estimates of Charness and Rabin (2002) and Bruhin, Fehr, and Schunk (2019).[35] According to our estimations, the large majority of subjects is of the *inequity-averse*, *competitive*, or *spiteful* type. Meanwhile, only 2.1% and 4.8% of subjects show a concern for *social welfare* or *altruism*, respectively, which are the most frequent preference types in the above studies.

Table 9 thus summarizes our social preference estimates from three alternative utility models and two estimates from a different subset of our experimental data. Column 1 replicates our estimates in Table 8, which are based on our preferred model (4) presented in Section 3.2. Columns 2–4 are based on a modification of this model. In particular, for Columns 2 and 3, we assume that subjects compare

[33]The poor performance of efficiency is not entirely surprising in the light of the experimental findings in Charness, Feri, Meléndez-Jiménez, and Sutter (2014). Efficiency concerns are particularly powerful in games where equilibrium outcomes can be Pareto ranked. This is not the case in our games with strategic substitutes.

[34]To see the intuition, start from a payoff-maximizing equilibrium profile in the dyad with $e_i + e_j = 12$. Suppose player i would mistakenly reduce her investment e_i by x. A best response by player j would then lead to the different equilibrium $(e_i - x) + (e_j + x) = 12$.

[35]The estimated distributions in Bellemare, Kröger, and Van Soest (2008) and Kerschbamer and Müller (2020) are in contrast closer to our preferred distribution.

Table 8: Preferred preference type distribution

Preference type	Preference strength: any (any $\hat{\epsilon}_s$)	moderate ($\hat{\epsilon}_s < 3$)	marginal ($\hat{\epsilon}_s < 1$)
altruism ($\hat{\rho}_s \geq \hat{\sigma}_s > 0$)	4.8%	2.1%	0%
social welfare ($\hat{\rho}_s > \hat{\sigma}_s = 0$)	2.1%	1.5%	1.0%
inequity-aversion ($\hat{\rho}_s > 0 > \hat{\sigma}_s$)	12.5%	3.3%	0.8%
competitive ($0 = \hat{\rho}_s > \hat{\sigma}_s$)	32.2%	25.1%	14.0%
spiteful ($0 > \hat{\rho}_s \geq \hat{\sigma}_s$)	23.6%	13.6%	3.1%
payoff maximizer ($\hat{\rho}_s = \hat{\sigma}_s = 0$)	23.4%	23.4%	23.4%
asocial ($\hat{\sigma}_s > 0 > \hat{\rho}_s$)	1.5%	1.5%	1.5%
Sum	100.0%	68.9%	42.4%

NOTES: Categorization of estimated $(\hat{\sigma}_s, \hat{\rho}_s)$-pairs according to revealed preference type and revealed preference strength. A value of $-0.0465 < x < 0.048$ for $x \in \{\hat{\sigma}_s, \hat{\rho}_s\}$ is set to zero because a player with such a small preference parameter would take a decision indistinguishable from a payoff-maximizer. The mapping of all other $(\hat{\sigma}_s, \hat{\rho}_s)$ into types and strengths is based on utility model (4) and Lemma 1 applied to the specific parameters of our experimental games (see Theoretical Appendix A.2 for this specification of Lemma 1).

their payoffs with all other players in the game. This is a plausible assumption to make because subjects can see everyone's payoffs on their screens.[36] For our estimations in Columns 2 and 4, we used, in turn, the investment decisions in the earlier rounds of our network games. These estimates thus alleviate the concern that the late-game decisions are "spoiled" by the earlier decisions of the other players. Finally, Columns 5 and 6 summarize our results when we estimate the original distributive preference model by Charness and Rabin (2002). According to this model, players' social preferences consist of the *difference* between their own payoffs and the payoffs of other players, rather than the absolute *level* of other players' payoffs as in model (4).[37] Again, just as in Columns 1 and 3, we estimated two variants of this model that differ in terms of how many players are included in a subject's reference group.

Overall, the preference estimates based on the final decision moments ($[30, t^{max}]$) are to be preferred over those from the early decision moments ($[20, 30)$). The latter seem to be heavily influenced by the initial conditions, that is, the fact that subjects click up from an initial zero investment. This at least is what the large shares of *spiteful* types in Columns 2 and 4 suggest. Moreover, the estimates in Column 1 are are preferable over those in Column 3 because the share of *asocial* types, with the counter-intuitive parameter constellation $\hat{\sigma}_s > 0 > \hat{\rho}_s$ is unreasonably large in Column 3.

For the same reason, we also prefer the estimates in Column 1 over those in Columns 5 and 6, which are based on the original Charness and Rabin (2002)

[36] Against this assumption speaks the fact that subjects only influence their neighbors' payoffs directly.

[37] In particular, we estimated the following model:

$$U_s(e_i, e_{-i}, i) \quad = \quad \pi_i \; + \; \frac{1}{|R_s|} \sum_{j \in R_s} \left(\rho_s \, r_{ij} + \sigma_s \, s_{ij}\right) (\pi_j - \pi_i) + \theta_{(e_i, t_s, i)},$$

where $\theta_{e_i, t_s, i}$ is a random utility component and

$$\begin{aligned} r_{ij} \quad &= \quad 1 \quad \text{if } \pi_i > \pi_j \text{ and } r_{ij} \; = \; 0 \text{ otherwise,} \\ s_{ij} \quad &= \quad 1 \quad \text{if } \pi_i < \pi_j \text{ and } s_{ij} \; = \; 0 \text{ otherwise.} \end{aligned}$$

Table 9: Alternative preference type distributions

Utility model:	Model (4)				C&R Model	
Reference group:	neighbors		all		neighbors	all
Estimation period:	$[30, t^{max}]$	$[20, 30)$	$[30, t^{max}]$	$[20, 30)$	$[30, t^{max}]$	$[30, t^{max}]$
Type distribution:	(1)	(2)	(3)	(4)	(5)	(6)
altruist	4.8	2.3	3.3	0.2	18.0	30.7
social welfare	2.1	1.2	2.0	1.3	4.4	5.4
inequity-averse	12.5	18.4	3.3	2.3	23.8	25.9
competitive	32.5	25.3	15.8	20.6	6.3	4.6
spiteful	23.6	38.4	15.6	38.1	14.4	5.6
money max	23.4	10.4	41.9	22.5	3.3	2.1
asocial	1.5	4.0	17.9	15.0	29.9	25.7
total	100	100	100	100	100	100

NOTES: Classification of estimated $(\hat{\sigma}_s, \hat{\rho}_s)$ into social preference types. Estimates are based on subject-and game-specific conditional logit estimations of six alternative utility models and/or investment choices from the beginning of the network games. See Section 4.4 for additional procedural details.

model. This model also shows an unreasonably high share of *asocial* types. An issue with this model in the context of the Bramoullé and Kranton (2007) game is that—depending on the curvature of the social benefit function $b(\cdot)$—the model can rationalize a downwards (or upwards) deviation from a payoff-maximizing best response as either an attempt to increase or to reduce payoff inequality. In particular, given the curvature of the benefit function implemented in our experiment, the model interprets the frequently observed downward deviations from the privately optimal level of $e^* = 12$ as attempts to *increase* payoff inequality when the focal player is *behind*. As a result, the model classifies many decisions as being motivated by a love of behindness, that is, a parameter estimate $\hat{\sigma}_s > 0$. Accordingly, it finds many *asocial* types, which are instead classified as *spiteful* or *competitive* types by our preferred model (4).

Nevertheless, it is interesting to note that the shares of *altruists* in Columns 5 and 6 are much in line with earlier estimates of the social preference type distribution that are based on decisions in dictator games (in particular Bruhin, Fehr, and Schunk, 2019). This suggests that our preferred estimates in Column 1 are not so much affected by the fact that we elicited them from the "heated" decision situations in our network games, where subjects take interactive decisions under time pressure. Rather, the original Charness and Rabin (2002) model and model (4) just seem to attach a different interpretation to the same choice.

B.4 Hypothesis 1: additional evidence

Our results should not be affected by the way we estimate the social preference types of our subjects. Intuitively, if the measurement affects all social preference estimates alike, then it should not change our conclusion regarding the compatibility of preferences in a subject group. In support of this view, Table 10 replicates our results on Hypothesis 1 based on the three most meaningful alternatives to our preferred social preference estimates. The results of all three models lend support to the key mechanism behind our theory: Groups with compatible preferences coordinate more likely on a refined ORE and less likely on a non-refined ORE than groups with incompatible preferences.

Table 10: Multinomial logit results based on alternative type distributions

	refined other-regarding eq.			non-refined other-regarding eq.		
	($\epsilon = 0$)	($0 < \epsilon < 3$)	($3 \leq \epsilon$)	($\epsilon = 0$)	($0 < \epsilon < 3$)	($3 \leq \epsilon$)
Model 1 (Type dis. 3):						
Pref. comp. (any $\hat{\epsilon}$)	0.373	0.285	base	-0.841	-0.768	-0.518
	(0.296)	(0.277)	outcome	(1.076)	(0.581)	(0.674)
Model 2 (Dis. 3, $t^{max} \geq 50$):						
Pref. comp. (any $\hat{\epsilon}$)	0.921**	0.574	–	-0.460	-0.688	-0.773
	(0.390)	(0.374)		(1.157)	(0.846)	(1.126)
Model 3 (Type dis. 5):						
Pref. comp. (any $\hat{\epsilon}$)	0.234	0.496*	–	0.145	-0.160	-0.664
	(0.286)	(0.269)		(0.617)	(0.427)	(0.591)
Model 4 (Dis. 5, $t^{max} \geq 50$):						
Pref. comp. (any $\hat{\epsilon}$)	0.318	0.611*	–	0.427	0.398	-0.504
	(0.378)	(0.360)		(0.737)	(0.532)	(0.824)
Model 5 (Type dis. 6):						
Pref. comp. (any $\hat{\epsilon}$)	0.263	0.443	–	-0.217	-0.110	-0.965
	(0.301)	(0.284)		(0.691)	(0.447)	(0.670)
Model 6 (Dis. 6, $t^{max} \geq 50$):						
Pref. comp. (any $\hat{\epsilon}$)	0.151	0.359	–	-0.423	0.193	-1.615
	(0.395)	(0.374)		(0.847)	(0.536)	(1.093)

NOTES: Models 1, 3, 5: 840 observations from final decision moments of all networks games except the games on the circle network. Models 2, 4, 6: 517 observations from final decision moments on or after the second-50 mark. All models include two group-specific *experience* measures (one measuring the position of a game in a session, the other measuring the x-th repetition of the same network game) and measures of network *size* and *clustering*. Standard errors clustered at the session level in parentheses: $^{***}p < 0.01$,** $p < 0.05$,* $p < 0.1$.

In the multinomial logit models of Table 3, the relationship between preference compatibility and equilibrium selection is strongest for those groups that feature at most a moderate or marginal social preference strength ($\hat{\epsilon} < 3$). These groups put most weight on the refined ORE with $\epsilon < 3$. This might raise the question whether it is truly the proper preference combination or rather just the absence of some pronounced social preferences in these groups that drives our findings. Obviously, a group of money maximizers would put most weight on a money-maximizing equilibrium profile, and so our preference *compatibility* measure might mistakenly pick up that effect. To exclude this possibility, we performed several additional tests where we directly tested the effect of preference compatibility in the subset of subject groups with small ($0 < \hat{\epsilon} < 3$) and strong social preferences ($3 \leq \hat{\epsilon}$). The results presented in Table 11 are by and large in line with what we saw in Table 3. Among the groups with a small social preference strength, the groups that put most weight on the refined ORE are the ones that have compatible preference.

Table 11: Multinomial logit results conditional on social preference strength

	refined other-regarding eq.			non-refined other-regarding eq.		
	$(\epsilon = 0)$	$(0 < \epsilon < 3)$	$(3 \leq \epsilon)$	$(\epsilon = 0)$	$(0 < \epsilon < 3)$	$(3 \leq \epsilon)$
Model 1 ($3 \leq \hat{\epsilon}$):						
Pref. comp.	0.749***	0.498*	1.029***	-0.261	0.011	base
	(0.207)	(0.294)	(0.287)	(0.402)	(0.373)	outcome
Model 2 ($0 < \hat{\epsilon} < 3$):						
Pref. comp.	0.936	1.173**	0.809	0.345	0.296	–
	(0.641)	(0.544)	(0.586)	(0.535)	(1.236)	–

NOTES: Observations from final decision moments of all network games, but the games on the circle: 532 (Model 1), 308 (Model 2). All models include two group-specific *experience* measures (one measuring the position of a game in a session, the other measuring the x-th repetition of the same network game) and measures of network *size* and *clustering*. Standard errors clustered at the session level in parentheses: $^{***}p < 0.01$,** $p < 0.05$,* $p < 0.1$.

B.5 Hypothesis 2: Additional evidence

Here, we present the regression results of our Placebo test on the circle network (Table 12) and the conditional probability table required for our quantitative predictions (Table 13).

Table 12: Placebo test for the circle network—Multinomial logit results

Compatible pref. of strength...	distributed/specialized eq. profiles ($\epsilon = 0$)	($0 < \epsilon < 3$)	($3 \leq \epsilon$)	other eq. profiles ($\epsilon = 0$)	($0 < \epsilon < 3$)	($3 \leq \epsilon$)
Model 1: any $\hat{\epsilon}$	-0.578	-0.885**	-0.885	2.196*	-0.346	base
	(0.593)	(0.374)	(0.679)	(1.276)	(0.397)	outcome
Model 2: $\hat{\epsilon} < 3$	-13.971***	-14.304***	-0.001	1.161**	-0.396	—
	(0.989)	(0.881)	(0.711)	(0.507)	(0.301)	—
Compatible pref. of strength...	distributed eq. profiles ($\epsilon = 0$)	($0 < \epsilon < 3$)	($3 \leq \epsilon$)	other eq. profiles ($\epsilon = 0$)	($0 < \epsilon < 3$)	($3 \leq \epsilon$)
Model 3: any $\hat{\epsilon}$	-0.125	13.607***	14.372***	—	-0.216	—
	(0.930)	(1.168)	(0.870)	—	(0.255)	—
Model 4: $\hat{\epsilon} < 3$	-0.949	-0.322	13.790***	—	-0.353	—
	(0.901)	(0.443)	(0.977)	—	(0.549)	—
Compatible pref. of strength...	specialized eq. profiles 1 ($\epsilon = 0$)	($0 < \epsilon < 3$)	($3 \leq \epsilon$)	other eq. profiles ($\epsilon = 0$)	($0 < \epsilon < 3$)	($3 \leq \epsilon$)
Model 5: any $\hat{\epsilon}$	-1.562	1.107	1.669	15.39***	-0.095	—
	(1.299)	(1.212)	(1.312)	(0.945)	(0.540)	—
Model 6: $\hat{\epsilon} < 3$	-13.989***	-14.041***	1.046	1.392**	-0.043	—
	(0.987)	(1.441)	(1.785)	(0.630)	(0.414)	—
Compatible pref. of strength...	specialized eq. profiles 2 ($\epsilon = 0$)	($0 < \epsilon < 3$)	($3 \leq \epsilon$)	other eq. profiles ($\epsilon = 0$)	($0 < \epsilon < 3$)	($3 \leq \epsilon$)
Model 7: any $\hat{\epsilon}$	15.735***	1.437	-1.090	0.729	-0.679*	—
	(0.747)	(1.134)	(0.792)	(1.502)	(0.360)	—
Model 8: $\hat{\epsilon} < 3$	-13.344***	-14.244***	0.753	0.591	-0.620	—
	(0.722)	(1.009)	(0.748)	(0.575)	(0.439)	—

NOTES: Results of eight multinomial logit estimations for the final decision moments in the circle network. Models 1–2: 360 observations, Models 4–8: 120 observations. The independent variable is an indicator variable that measures whether a subject group has a preference combination that matches the compatibility requirements for the complete (Models 3–4) or the star (Models 5–8) or either of the two networks (Models 1–2). Concretely, for Models 5–8, we searched for groups where every second subject matches the compatibility requirements for the star center and every other subject matches the requirements for the star periphery. The dependent variable in Models 3–4 (Models 5–8) is a multinomial variable that tests whether a subject group plays a distributed (one of the two specialized) profile in the circle. Models 1–2 test both together. All models include two group-specific *experience* measures (one measuring the position of a game in a session, the other measuring the x-th repetition of the game on the circle network). Standard errors clustered at the session level in parentheses: $^{***}p < 0.01$,$^{**}p < 0.05$,$^{*}p < 0.1$.

Table 13: Frequencies of predicted ORE profiles conditional on group social preference strength

		Deviation from payoff-maximizing equilibrium								
		$\epsilon = 0$			$\epsilon < 3$			any ϵ		
		Group with preference strength								
Network	Equilibrium	$\hat{\epsilon} < 1$	$1 \leq \hat{\epsilon} < 3$	$3 \leq \hat{\epsilon}$	$\hat{\epsilon} < 1$	$1 \leq \hat{\epsilon} < 3$	$3 \leq \hat{\epsilon}$	$\hat{\epsilon} < 1$	$1 \leq \hat{\epsilon} < 3$	$3 \leq \hat{\epsilon}$
Dyad	egalitarian	7.7%	1.5%	0.3%	7.7%	4.6%	1.0%	7.7%	4.6%	5.9%
	other	92.3%	18.5%	4.2%	92.3%	95.4%	21.5%	92.3%	95.4%	94.1%
Complete	egalitarian	0.2%	0.04%	0%	0.2%	0.1%	0%	0.2%	0.1%	0.02%
	other	99.8%	19.4%	0.5%	99.8%	99.9%	2.8%	99.8%	99.9%	100%
Star	per-sp, cent-sp: $\pi_c \geq \pi_j$	50.0%	0.0%	0%	50.0%	0.8%	0.0%	50.0%	0.8%	85.7%
	cent-sp: rest	50.0%	0.1%	0%	50.0%	99.2%	3.9%	50.0%	99.2%	14.3%
Circle	specialized	66.7%	0.3%	0%	66.7%	2.8%	0.02%	66.7%	2.8%	0.3%
	distributed	33.3%	1.6%	0%	33.3%	97.2%	0.7%	33.3%	97.2%	99.7%
Core	per-sp, cent-sp: $\pi_c \geq \pi_j$	92.9%	0.4%	0.02%	92.9%	5.6%	0.2%	92.9%	5.6%	83.3%
	cent-sp: rest	7.1%	0.0%	0%	7.1%	94.4%	3.9%	7.1%	94.4%	16.7%
D-box	per-sp, cent-sp: $\pi_c \geq \pi_j$	7.1%	0.0%	0%	7.1%	2.0%	0.1%	7.1%	2.0%	59.4%
	cent-sp: rest	92.9%	0.1%	0.02%	92.9%	98.0%	5.1%	92.9%	98.0%	40.6%
Line	end-sp, distr: $\pi_c \geq \pi_j$	46.2%	0.1%	0.01%	46.2%	92.9%	5.2%	46.2%	92.9%	42.6%
	distr: rest	53.8%	0%	0.01%	53.8%	7.1%	0.4%	53.8%	7.1%	57.4%

NOTES: Numbers are calculated as % of the total number of possible investment profiles for a group of players with a social preference strength of $\hat{\epsilon} < 1$, $1 \leq \hat{\epsilon} < 3$, and $3 \leq \hat{\epsilon}$, respectively.

C Replication instructions

C.1 Experimental design

The computerized experiment was designed using the software program z-tree 3.0 (Fischbacher, 2007) and conducted in the Experimental Laboratory for Sociology and Economics (ELSE) at Utrecht University between 09.06. and 18.06.2008.

In the experiment, subjects had to invest in the production of a local public good in each of the seven network structures shown in Figure 1. In total, eight experimental sessions of approximately one-and-a-half hours length were scheduled and completed. Using the ORSEE recruitment system (Greiner, 2004), over 1,000 potential subjects were approached by e-mail to participate in the experiment. On average, 15 students participated per session, which gives 120 subjects in eight sessions.

A session consisted of seven treatments with varying order of treatments between the sessions. Each network structure represents a different treatment. Table 9 gives an overview.

Session	Ordering	Treatment						
		1	2	3	4	5	6	7
1	1	Dyads	Line	Star	Square	Core	Dbox	Complete
2	2	Complete	Dbox	Core	Square	Star	Line	Dyads
3	3	Dyads	Star	Line	Core	Square	Dbox	Complete
4	4	Complete	Dbox	Square	Core	Line	Star	Dyads
5	3	Dyads	Star	Line	Core	Square	Dbox	Complete
6	2	Complete	Dbox	Core	Square	Star	Line	Dyads
7	1	Dyads	Line	Star	Square	Core	Dbox	Complete
8	4	Complete	Dbox	Square	Core	Line	Star	Dyads

Figure 9: Order of treatments by session

General instructions were given before the start of a session (see the instructions below). In each treatment, subjects played a local public goods game on a given network structure fives times, for 60 seconds on average under the same conditions. In particular, being positioned in a specific network, subjects could invest for a limited amount of time in order to improve their experimental points that were calculated based on formula (2).

The five repetitions of a treatment are called rounds, and each treatment consisted of one trial round and four payment rounds. At the beginning of a round, subjects were randomly allocated to a group together with either one or three other participants. Subjects were indicated as circles on the screen and could identify themselves by color: Each subject saw him- or herself as a blue circle while all neighbors were represented as black circles (see below for a screenshot).

Each round had the same structure and lasted between 30 and 90 seconds. The round ends were unknown and randomly determined. Starting from a situation of zero investments, subjects indicated simultaneously on their computer terminals (by clicking on one of two buttons at the bottom of the screen) whether they wished to in- or decrease their investment. Full information about the momentary investments of all other subjects was continuously provided and updated five times per second by the computer. Also, the resulting payoffs of all participants could continuously be observed on the screen. At the end of a round, subjects were informed about the number of points they earned with the investments they saw on their screen. In other words, final earnings only depended on the situation at the end of a round.

Subjects were not identifiable between different rounds or at the end of the experiment. In this fashion, we minimized the dependence across observations (Falk and Kosfeld, 2003). Taking the seven treatments together, every subject played 35 network games in 35 different groups, of which 28 were payoff relevant. Altogether, this gives 960 networks games and 3,360 investment decisions (8 sessions

times 15 subjects on average per session times 7 treatments (6 networks of 4 subjects and 1 network of 2 subjects) times 4 cycles). At the end of the experiment, the experimental points were converted into euros at a rate of 400 points = 1 euro. In addition, subjects received a 3 euro participation fee. The average earning was thus 11.82 euros.

C.2 Experimental instructions

C.2.1 English version

- 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 -

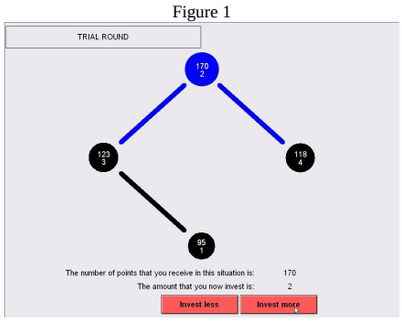
The experiment consists of *seven scenarios*. Each scenario consists again of *one trial round* and *four paid rounds* (altogether 35 rounds of which 28 are relevant for your earnings).

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



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.

C.2.2 Dutch version

- Instructions -

Neemt u alstublieft de volgende instructies aandachtig door. Hierin staat alles wat u moet weten om deel te nemen aan het experiment. Indien u vragen hebt, steek uw hand op. Er zal iemand bij u komen om uw vraag te beantwoorden. Deze regels zijn hetzelfde voor alle deelnemers.

U kunt geld verdienen tijdens dit experiment door het vergaren van punten. Het aantal punten dat u verdient, hangt af van uw eigen keuzes en van de keuzes van andere deelnemers. Het totaal aantal punten dat u verdient in het experiment zal aan het einde van het experiment omgewisseld worden tegen de wisselkoers van:

400 punten = 1 Euro

Aan het einde van het experiment krijgt u het geld dat u verdiend hebt tijdens het experiment contant uitbetaald. Later volgen hierover verdere instructies. Tijdens het experiment is het niet toegestaan te communiceren met andere deelnemers. Zet u mobiele telefoon uit en berg hem op in uw tas. U mag ook alleen de functies op het scherm activeren die nodig zijn voor het functioneren van het experiment. Hartelijk dank.

- Overzicht van het experiment -

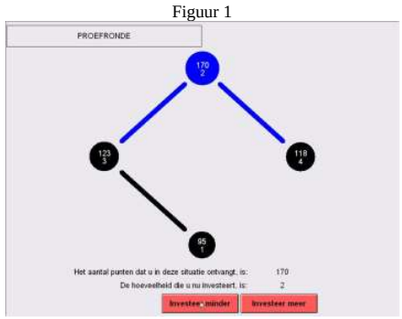
Het experiment bestaat uit *zeven scenario's*. Elk scenario bestaat weer uit *één proefronde* en *vier betaalde rondes* (samen 35 rondes waarvan er 28 relevant zijn voor uw verdiensten).

In alle scenario's wordt in een groep geplaatst met één of drie andere deelnemers. Aan het begin van *elk van de 35 rondes* worden de groepen en de posities binnen de groepen willekeurig veranderd. De deelnemers waarmee u in de ene ronde in een groep zit, zijn zeer waarschijnlijk andere deelnemers dan diegene waarmee u in de volgende ronde in een groep zit. Tijdens of na het experiment zal het niet bekend worden gemaakt met wie u in een groep gezeten hebt.

De deelnemers in uw groep (dat zijn er twee of vier afhankelijk van het scenario) worden als cirkels weergegeven op het scherm (zie Figuur 1). U wordt zelf weergegeven met een **blauwe** cirkel, terwijl de andere deelnemers worden weergegeven als **zwarte** cirkels. U bent altijd verbonden met één of meer andere deelnemers in uw groep. Deze andere deelnemers noemen we *uw buren*. Deze verbindingen verschillen per scenario and worden weergegeven met lijnen tussen de cirkels op het scherm (zie ook Figuur 1).

Elke ronde duurt *tussen de 30 en 90 seconden*. Het einde zal op een onbekend en willekeurig moment in dit tijdsinterval plaatsvinden. Tijdens dit tijdsinterval kunt u punten verdienen door kennis te produceren, maar de productie van kennis kost ook punten. De punten die u aan het einde ontvangt, hangen af van uw eigen investering in kennis en van de investeringen van uw buren.

Figuur 1



Door te klikken op de twee knoppen onder aan het scherm, kunt u uw investering in kennis verhogen of verlagen. Aan het einde van elke ronde, ontvangt u het aantal punten dat op dat moment op het scherm wordt weergegeven. Uw uitbetaling hangt dus alleen af van de situatie aan het einde van elke ronde. Merk op dat dit einde komt op een voor iedereen onbekend moment tussen de 30 en 90 seconden na het begin van de ronde. Verschillende rondes zullen ook niet even lang duren.

Het aantal punten dat u zult ontvangen zijn weergegeven als het bovenste getal in uw blauwe cirkel. De punten die anderen zullen ontvangen zijn weergegeven als het bovenste getal in hun zwarte cirkels. Daarnaast verandert *de grootte van de cirkels* met het aantal punten dat u of de andere deelnemers zullen krijgen: een groter cirkel betekent dat die deelnemer meer punten zal verdienen. Het onderste getal in de cirkels geeft het aantal punten weer dat de deelnemers in uw groep investeren in kennis.

Opmerkingen:

- Het kan gebeuren dat er een vertraging is tussen uw klik en de veranderingen van de getallen op het scherm. Eén klik is voldoende om uw investering met één punt te veranderen. Een volgende klik zal pas effect hebben als de eerste klik is verwerkt.
- **Wacht daarom met een volgende klik totdat uw eerdere verandering verwerkt is op het scherm!**

- Uw verdiensten -

Nu leggen we uit hoe uw verdiensten afhangen van de investeringen. Lees dit zorgvuldig! Wees niet bezorgd als het niet meteen helemaal duidelijk is. We zullen zodadelijk ook een rekenvoorbeeld laten zijn. Daarnaast is er bij elk scenario een proefronde om ervaring te krijgen met hoe uw investering uw aantal punten bepaalt.

In alle scenario's hangt het aantal punten dat u ontvangt aan het einde van een ronde af van twee factoren:

1. **Elke eenheid die u investeert in kennis kost uzelf 5 punten**.
2. **U verdient punten met elke eenheid die uzelf investeert en met elke eenheid die uw buren investeren.**

Als u de hoeveelheid die uzelf investeert en de investeringen van uw buren optelt, geeft de volgende tabel weer hoeveel punten u verdient met deze investeringen:

Uw investeringen plus investeringen van uw buren	0	1	2	3	4	5	6	7	8	9	10
Punten	0	28	54	78	100	120	138	154	168	180	190

Uw investeringen plus investeringen van uw buren	11	12	13	14	15	16	17	18	19	20	21
Punten	198	204	208	210	211	212	213	214	215	216	217

Hoe hoger de investeringen worden, hoe minder punten erbij komen voor nieuwe investeringen. Als het totaal van investeringen boven de 21 komt, ontvangt u nog één punt voor elke volgende eenheid die u of een van uw buren investeren.

Let op: als uw investering plus die van uw buren samen 12 of meer zijn, stijgen uw verdiensten met minder dan 5 punten per extra eenheid investering.

- Voorbeeld -

Stel

1. u investeert 2 eenheden;
2. één van uw buren investeert 3 eenheden, een andere buur 4 eenheden.

Dan moet u 2 keer 5 = 10 punten betalen voor uw eigen investering.

De investeringen waarvan u profiteert zijn uw eigen investering plus die van uw buren: 2 + 3 + 4 = 9. In the tabel kunt u zien dat dit u 180 punten oplevert.

In totaal betekent dit dat u 180 − 10 = 170 punten verdient als dit de situatie zou zijn aan het einde van de ronde. Figuur 1 laat dit voorbeeld zien zoals het op uw scherm verschijnt. De investering van de vierde deelnemer in uw groep heeft geen effect op uw aantal punten. In de proefronde aan het begin van elk scenario krijgt u de kans om te wennen aan hoe de punten die u ontvangt veranderen met de investeringen.

- Scenario's -

Alle rondes zijn in principe hetzelfde. Het enige wat verandert tussen de scenario's is de manier waarop u met andere deelnemers verbonden bent. Ook zal uw eigen positie in een groep kunnen veranderen tussen rondes. U krijgt elke keer een mededeling op het scherm als een nieuw scenario en een proefronde begint. Bovenaan het scherm kunt u ook zien of u in een proefronde zit (zie Figuur 1). Betaalde rondes worden aangegeven met alleen "RONDE", terwijl proefrondes worden aangegeven met "PROEFRONDE".

- Vragenlijst -

Aan het einde van de 35 rondes vragen we u nog om een vragenlijst in te vullen. Neem alstublieft rustig te tijd om deze vragenlijst precies in te vullen. Ondertussen tellen wij uw verdiensten. Blijft u op uw plek totdat de betaling is afgerond.

3

C.3 Selection and eligibility of participants

Subjects subscribe to a database via a website (www.elseutrecht.nl), which explains the type of experiments that they subscribe for. The Welcome-text is shown in Appendix C.4.

All subjects are recruited from this database. By subscribing a subject indicates her willingness to participate in the type of experiments described. This means that by subscribing, a subject in principle agrees to participate in the described type of task. All experiments exclusively involve computerized tasks.

At the beginning of an experiment, subjects are informed that if for any reason they might not be willing to continue, they can notify the experiment leader and stop the experiment (for details on the rules, see elseutrecht.nl/public/rules.php). No further explicit consent form is used for individual experiments.

C.4 Recruitment text

Recruitment text participants

Welcome! This is the web site of the "Experimental Laboratory for Sociology and Economics"(ELSE) at Utrecht University.

ELSE is a computer room. It is used for studying social science and economics research questions in an experimental manner. For this purpose, we are looking for people who are interested to participate in our experiments. During an experiment, up to 30 participants at a time can get a place at one of the computers in ELSE. The participants anonymously interact with other participants via the computer and answer a short questionnaire about themselves. On average, participants earn between 8 and 10 EURO per hour, but this amount can vary between studies. The exact amount of money received typically depends on the decisions made by oneself and other participants. The experiments do not involve other tasks than making decisions during anonymous interactions, and answering questions via the computer, unless this is explicitly mentioned in an invitation.

Are you interested in earning money for your decisions and answering our questions, support science and gain some insight into this research field? Then be welcome to participate in our experiments!

In order to participate, you first need to subscribe to the participants database via this web site (click on register in the menu on the left side). Every now and then you will receive a message inviting you for a specific experiment. At that moment you can decide whether you want to participate in a specific experiment (see Rules)

Before you subscribe to the participants database, you need to indicate that you agree with the rules we follow. Please read the information provided on our web site concerning:

- Participation and Rules of Proper Laboratory Behavior -
Researchers' Commitments and Privacy Policy -
Frequently Asked Questions (FAQs)

If you have further questions, please feel free to contact us using the e-mail address below.

ELSE is located at the Uithof in the Sjoerd Groenmangebouw, Padualaan 14 at the 3rd floor in room A3.03.

C.5 Approval of the Institutional Review Board

This experiment is one of a series of experiments conducted for a project entitled "Cooperation in Social and Economical relationship". The Ethics committee of Social and Behavioral Sciences of Utrecht University granted joint approval to all the experiments of this project including the current experiment. The approval was filed on October 22, 2017, under number FETC17-028.

Universiteit Utrecht

P.O. Box 80140, 3508 TC Utrecht

The Board of the Faculty of Social and Behavioural Sciences
Utrecht University
P.O. Box 80.140
3508 TC Utrecht

Faculty of Social and Behavioural Sciences
Faculty Support Office
Ethics Committee

Visiting Address
Padualaan 14
3584 CH Utrecht

Our Description	FETC17-028 (Buskens)
Telephone	030 253 46 33
E-mail	FETC-fsw@uu.nl
Website	https://intranet.uu.nl/facultaire-ethische-toetsingscommissie-fetc
Date	October 22, 2017
Subject	Ethical approval

ETHICAL APPROVAL

Study: cooperation in social and economical relations

Principal investigator: prof. Vincent Buskens, Ph.D.

This research project does not belong to the regimen of the Dutch Act on Medical Research Involving Human Subjects, and therefore there is no need for approval of a Medical Ethics Committee.

The study is approved by the Ethics Committee of the Faculty of Social and Behavioural Sciences of Utrecht University. The approval is based on the documents send by the researchers as requested in the form of the Ethics committee and filed under number FETC17-028 (Buskens). Given the review reference of the Ethics Committee, there are no objections to execution of the proposed research project, as described in the protocol. It should be noticed that any changes in the research design oblige a renewed review by the Ethics Committee.

Yours sincerely,

Peter van der Heijden, Ph.D.
Chair

Jacqueline Tenkink-de Jong LLM
Executive secretary