Purdue University Economics Department Working Paper No **1323**

The Network Dilemma*

Evan M. Calford[†] and Anujit Chakraborty[‡]

February 12, 2020

Abstract

This paper introduces and studies the Network Dilemma, which embeds a social dilemma within a network formation game. The game models the voluntary provision of a costly public good that is *locally* non-excludable within an endogenously formed network of agents. In this environment, all equilibrium networks are star networks that exhibit ex-post payoff inequality between the central public good provider and peripheral players. In a laboratory experiment, we observe high levels of equilibrium behavior in the repeated simultaneous game, despite the coordination problem faced by the ex-ante homogeneous subjects. In a sequential version of the Network Dilemma we observe high levels of efficient non-subgame perfect equilibrium behavior, where the first mover "volunteers" to provide the public good. Using a diagnostic treatment we find that aversion to behavioral uncertainty, and not other-regarding preferences, is the main motivating factor behind voluntary public good provision observed in the Network Dilemma.

Keywords: Network formation, Volunteer's dilemma, experimental economics, coordination games

JEL codes: D85, D91,D81, C92

*We thank seminar audiences at UCL, UC Davis, Purdue University, IIMA, Academia Sinica, and the ESA World Meeting. Conversations with Rod Garratt, Burkhard Schipper and Ala Avoyan are gratefully acknowledged. We thank Phillip Chapkovski for providing advice and sample code for software implementation. Chen Wei and William Brown provided very helpful research assistance. Funding for subject payments was provided by the Krannert School of Management, Purdue University.

[†]Krannert School of Management, Purdue University; ecalford@purdue.edu.

[‡]Department of Economics, University of California, Davis; chakraborty@ucdavis.edu.

1 Introduction

Local public goods games are an important testing bed for understanding a range of economic phenomena that create *local* externalities: experimentation in the face of market shocks (Guiso and Schivardi, 2007), knowledge spillover from R&D (Jaffe et al., 1993), experimentation with agricultural input ratios (Foster and Rosenzweig, 1995), and the distribution of product information via "market mavens" (Feick and Price, 1987) for example.

To illustrate, consider the stylized example of technological innovation and transmission across four firms owned by four different countries, who are potential trade-partners and who have monopoly-power in their national market. State ownership could be understood as a simplification of the national governments having significant stakes in firm performance (e.g, via lobbying, or because the firms are partly state-owned, or because the employment provided by the firm is politically valuable for incumbent politicians). Each country may either innovate themselves, or attempt a technology transmission through investing in mimicking innovations from a foreign firm, and the decisions are taken simultaneously. Transmission of innovation from one country to another only happens if resources were spent by the latter country on mimicking the former country and the former country has invested in innovation. Thus, conditional on there being an innovator, the externality from her innovation is (locally) propagated to the mimickers of this particular innovator. Mimicking could be achieved through technology stealing[1] or corporate espionage[2], or by allowing the innovating firm to create a trade-partnership in the local country followed by technology transfer. Access to new innovation is costlier if the access was acquired by R&D investment, and cheaper if obtained by mimicking an innovating firm through one of the channels mentioned above. Despite this, innovation under costly R&D investment is a better outcome than spending resources on mimicking a firm that did not themselves innovate. This constitutes a game of technology innovation and sharing with local externalities, where countries face a dilemma about which role

[1] In January 2020, the head of Harvard's chemistry and chemical biology department, Charles Lieber, was swept up in one of the near two hundred cases of US government's investigations into scientists stealing research for other countries. According to an affidavit supporting the criminal complaint against Lieber, his contract with a Chinese program afforded him a monthly salary of up to $50,000, annual living expenses of more than $150,000 over three years, and more than $1.5 million by the Chinese government and the Wuhan University of Technology (WUT). In February 2020 he was reportedly arrested and charged with allegedly lying about his links to the Chinese university (Subbaraman, 2020).

[2] Micron accused Taiwan based company UMC and Fujian Jinhua Integrated Circuit, a state-backed chip maker from China, of technology theft in documents filed last December in Federal District Court for the Northern District of California, as reported in the NY Times (Mozur, 2018). According to the report, UMC first lured Micron's secrets from engineers previously working at Micron's Taiwan operations with promises of raises and bonuses.

(innovator vs mimicker) they would like to play and, in case they opt out of innovating, who to mimick.

Outside of economics, one could also consider the breeding habits of the Common Goldeneye duck (and other brood parasites) as an example of similar dynamics. A Goldeneye prefers to lay its eggs in another Goldeneye nest, conditional on the other bird raising the offspring, to avoid the costs associated with parenting. The next best option is for the Goldeneye to raise the eggs found in its own nest (including its own, and possibly others' eggs), thus providing a local externality to only the birds who have laid eggs at her nest. Laying eggs in the nest of a bird who does not care for the chicks leads to the worst evolutionary outcome: no offspring.[3]

The local public good game we introduce to model such economic environments is a network formation game (henceforth called the Network Dilemma or ND game) in which each player may provide a discrete public good or form a link to a potential public good provider. The benefits from the public good are accrued as long as you, or someone you are directly connected to in the network, has provided the public good, but public good provision is costly. Holding the behavior of others constant, a player's pecuniary benefit is maximized by linking to a player who is providing the public good. If a player provides the public good herself then she earns an intermediate pecuniary benefit. The lowest pecuniary benefit accrues to a player who does not provide the public good herself and has not managed to directly link to a player who provides the public good. This game constitutes a social dilemma as the costs of linking to other players are smaller than that of providing the public good, and any person providing the public good also provides externalities to all directly connected neighbors. One could conceive of the ND as a hybrid of a volunteer's dilemma and a network-mediated coordination game – an efficient (or pure strategy equilibrium) outcome requires exactly one player to provide the public good, and thus receive a smaller income in the efficient equilibrium than the others, but also requires every other player to know the identity of (and link to) the public good provider. In the context of our leading example, the efficient outcome requires one firm to innovate through investment in R&D, and for all other firms to benefit from the innovator in their respective jurisdictions, either through agreeable trade relationships or through

[3]The case of agricultural experimentation in developing countries, that is often mimicked by neighboring farmers, is also particularly relevant to our model.

some of the less ethical practices mentioned above.[4,5]

Motivated by the past literature on public good provision and behavior in networks, we organize our hypotheses around two plausible behavioral forces: other-regarding preferences and the strategic considerations of rational individuals (modeled through both equilibrium and rationalizability in this paper). A firm/country that dislikes disadvantageous profit (payoff) inequality may not be willing to provide innovation in the face of disagreeable mimicking of technology, whereas a country that primarily cares about its legacy in technological progress (impure altruism) or only cares about absolute profit, might readily volunteer as an innovator nonetheless.[6] Additionally, we also consider a particular kind of behavioral uncertainty (in first and higher order beliefs) as a potential motivation for public good provision: a country that is uncertain about motives of the other countries (e.g, their degree of inequality aversion) may choose to provide innovation itself. Uncertainty about the potential trade-partner's preferences ("dislikes disadvantageous inequality" versus "cares about its legacy in technological progress") generates uncertainty about the trade-partner's innovation propensity and hence about one's own profits/payoffs when depending on her provision, while personally investing in the innovation ensures a safe moderate outcome. Further, even if all countries have correct and degenerate (first-order) beliefs about their trade-partner's preferences, uncertainty about higher order beliefs may affect the final outcome of the game. Country i's belief over "country j's beliefs over the preferences of country k", affects country i's beliefs over the *strategy* of country j, and therefore affects country i's decision regarding R&D investments.

In the simultaneous version of the game, all actions are rationalizable, and hence exclusively studying the simultaneous ND game provides little insight into the nature of the motivations that are responsible for observed actions. To separate behavior consistent with common knowledge of selfish

[4]For example, when a foreign company wants to enter the Chinese market, it often has to surrender its technology to Chinese companies through a joint venture agreement. There are about 300,000-plus foreign-local joint ventures in China. Sometimes this technology transfer is coerced. Bloomberg (2019) reported "The European Union Chamber of Commerce in China found that 20% of companies surveyed in 2019 felt compelled to hand over know-how to maintain market access, up from 10% in 2017." The South China Morning Post (Cai and Elmer, 2019) reported "This pressure was particularly evident in hi-tech industries, with 44 per cent of aerospace and 41 per cent of chemical companies having reported facing "notable" pressure to transfer technology."

[5]For the Common Goldeneye, given the fixed costs of building and maintaining a nest, the efficient outcome within a small local community of birds is for one bird to raise the eggs in her nest and for other birds to lay eggs in that nest.

[6]Until recently the United States has regularly voluntarliy entered into trade relationships, including technology transfers, that might be characterized as having an altruistic component. President Trump, however, has shown a strong determination to avoid trade relationships that benefit foreign countries more than the United States and has frequently expressed a desire for "winning" trade deals.

preferences from that consistent with behavioral uncertainty over more general motivations and beliefs of others, we implement a four player sequential version of the ND with perfect information, where players are exogenously assigned to a move order. In the context of our technology-transfer example, suppose that some firms can pre-commit to investing in R&D themselves, or, unilaterally investing in institutions and arrangements that facilitate technology transfer.[7] The firms moving later can react according to the action taken by the previous firms. The theoretically-motivated assumption of a fixed, exogenous, move order provides us a platform to diagnose and understand the motivations of behavior within the ND structure.

Under common knowledge of selfish preferences, early movers have a strategic advantage in the sequential ND: each country may invest in institutions and arrangements that facilitate technology transfer from the country moving last. The final mover then maximizes their payoff by undertaking R&D, thereby providing the public good to the entire network. Consequently, public good provision by the first mover (P1) is not consistent with any model that assumes common knowledge of self-interested payoff-maximization (proposition 3). But when such a common knowledge assumption is realistically relaxed, countries that move earlier face more uncertainty about the beliefs and motivations (for example, selfish vs inequality averse vs impurely altruistic) of others: For example, the last mover (P4) observes all preceding moves and therefore does not face any uncertainty, the second to last mover (P3) faces first-order uncertainty about the motivations of P4, the third to last mover (P2) faces first-order uncertainty about P3 and P4's motivations and second-order uncertainty about the beliefs of P3 about the motivations of P4, and so on. As suggested by the intuition above, P1 volunteering to provide the public good can be accommodated by a model that allows expected payoff maximization under first- and higher-order uncertainty over the preferences of others (proposition 7). When implementing the sequential ND in the laboratory *we observe P1 volunteering to provide the public good as the modal outcome in all treatments*, and we rarely observe P4 providing the public good. Our data, therefore, rejects common knowledge of selfish preferences.

Our results cannot, however, be explained by an aversion to payoff inequality. By design, the final player to move faces no uncertainty about the behavior of earlier movers. The behavior

[7] Bloomberg reports, "...the most egregious cases (of forced technology transfer) are alleged to flow from environmental-impact assessments or other steps required for an operating license. Take chemicals, for example. In the U.S., inspections for new chemical plants may require reporting a range of temperatures at which a manufacturing process operates, whereas in China the specific temperature must be disclosed, according to Jacob Parker, vice president for operations at the U.S.-China Business Council in Beijing. 'Every new detail that's required allows for the replication of the process by potential domestic competitors,' he says."

of the fourth player (P4) therefore allows us to cleanly identify the preferences of P4. We find limited evidence for inequality averse behavior except in the treatment where payoff inequality in the efficient outcome is highest (12 points for the public good provider, and 30 points for other players). Even in this high payoff inequality treatment, the level of inequality averse behavior observed is not sufficient to justify the observed behavior of P1 volunteering to provide the public good (under an assumption of common knowledge of preferences).

We are left with two competing explanations for the behavior of P1. Either behavioral uncertainty faced by P1 about the behavior of later moving players, or that P1 simply *prefers* to be the public good provider (because of, for example, impure altruism preferences).

We are able to attribute the provision of the public good by P1 to her reaction to behavioral uncertainty, and rule out impure altruism as a candidate explanation. To test these two competing explanations, we designed a diagnostic sequential treatment that reduces the uncertainty experienced by P1 while leaving the payoffs unchanged from the baseline ND. In this diagnostic sequential treatment, if all but one of the players link to the remaining player then the remaining player's action is "reset" to providing the public good. This simple modification substantially reduces the extent of behavioral uncertainty experienced by the players (especially P1), even while allowing for preferences and beliefs over preferences to contain selfish traits, aversion to inequality, or impure altruism. It is now common knowledge that if any three players have connected to a fourth player, P*, then by design P* will be compelled to provide the public good. Folding a step back, in the case where any two players have already linked to P*, the remaining non-P* player has only one way to guarantee the maximum payoff for her and an efficient outcome for everyone else: by connecting to P*. Further, the first two players connecting to P* know these two facts about the other two players, and they each know that the other player knows it too. Hence there is no uncertainty about behavior on this path. When we implement the modified ND in the laboratory, we observe a remarkable shift in behavior when compared to the baseline sequential ND. *In the modified ND, the modal outcome is for the first three movers to link to P4 and for P4 to provide the public good.* In contrast, recall that in the baseline sequential ND the modal outcome is for P1 to volunteer immediately to provide the public good.

The ability to control and induce payoff functions and information flows in an experimental laboratory has recently led to the lab becoming a fertile testing ground for works on the provision of local public goods (Rong and Houser (2015), Goyal et al. (2017), van Leeuwen et al. (Forthcoming)). These papers have primarily focused on the question of when and how much local public good is provided. In this paper we answer a complementary question: who provides the public good when observable payoffs are ex-ante symmetric, and what behavioral forces drive this behavior.

Network experiments have the following two distinguishing features: i) The genesis of the network, whether it is exogenously assigned (Rosenkranz and Weitzel (2012); Charness et al. (2014); Mukherjee (2019); Carpenter et al. (2012) for example) or endogenously formed by individual decisions (Falk and Kosfeld (2012); Goeree et al. (2009) and this paper for example), and, ii) The structure of the network-mediated externalities that are included in the environment. In our design subjects may either form a link to another player or provide a binary public good (rather than both forming links, and providing non-binary public goods). Links provide only one-way transmission of public good benefits (towards the sponsor of the link)[8], and the transmission of benefits only accrue to immediate neighbours. Therefore, our subjects face a simplified choice environment that allows us to identify the effects of public good externalities independently of other network externalities that may exist in a more general environment.

Charness et al. (2014) study the effects of network uncertainty on behavior in the case of both strategic complements and strategic substitutes. In their paper, uncertainty over the network structure is exogenously induced by the experimenter, whereas in the current paper uncertainty over the network structure is endogenous and generated from uncertainty over the behavior of other subjects. Eichberger and Kelsey (2002) also study strategic uncertainty in public goods games. The focus in Eichberger and Kelsey (2002) is on the inter-relation between ambiguity aversion and strategic uncertainty, and they conclude that strategic uncertainty should lead to increased levels of public good provision. Although the mechanism and source of uncertainty considered in this paper is distinct from that studied by Eichberger and Kelsey, our experimental results are consistent with their conjecture: we find that subjects are more willing to volunteer to provide a public good if they perceive greater uncertainty about the behavior of others. Second (and higher) order beliefs are crucial to the results obtained in our paper but are not studied in the prior literature on uncertainty regarding the provision of public goods.

Section 2 describes the ND game and provides the theoretical results that form the backbone of the hypotheses that we explicitly state in Section 3. The results are presented in Section 4. Section 5 introduces the diagnostic treatment that we use to tease out the behavioral hypothesis that best explains the trends in our data. In Section 6, we provide an overview of the related literature, and Section 7 concludes.

[8]We relax this in some treatments.

2 The Network Dilemma

In general, a network formation game with public good provision exhibits three tightly connected sources of externalities.[9]

Externality 1 (Network-mediated Public good externality): Keeping the network structure fixed, changing the amount of public goods provided by any player changes the final payoffs of connected players.

Externality 2 (Reflexive externality of unilaterally formed bidirectional links): Keeping the network and amount of public goods provided by every player constant, any new bi-directional link (benefits flow in both directions) formed by one of the parties (assuming such links can be formed unilaterally), potentially affects the non-consenting player's utility by creating new paths in the network involving the non-consenting player.

Externality 3 (Third party Externality): Keeping the network and amount of public goods provided by every player constant, a new link made between any 2 players, creates potential benefits downstream (in the direction of the benefit-flow) for third-party players.

In the ND we isolate only the first externality, by making the flow of public good benefits unidirectional towards the party investing in the link (ruling out Externality 2), and assuming that public good benefits only create externalities for immediate neighbors (ruling out Externality 3).[10] Thus the ND game is a game of purely *Network-mediated Public good externality.*

We consider the four player, five node, Network Dilemma where each player resides on a single node. We refer to the node that a player resides on, and the player residing on the node, interchangeably using the notation Pi for $i \in \{1, 2, 3, 4\}$. The empty node is denoted by $P0$, and this is the the *source of the public good.* Each player must form exactly one directed edge to a node other than the node they occupy.

In our study, providing the public good is equivalent to connecting to $P0$. This action earns $B \in \{12, 24\}$ points, which is the net profit from supplying the public good benefit while paying the cost of provision. If a player links to another player, Pj, then they earn 30 points if Pj links to $P0$ (the net profit of public good access, while paying the linking cost to another player) and 10 points otherwise (a non-zero baseline profit). Consistent with our motivating examples, the highest payoff is attained by free-riding successfully on someone who is providing the public good, the lowest is

[9]Galeotti and Goyal (2010) presents a prototypical example. See Section 6 for further examples and discussion.

[10]In Galeotti and Goyal (2010), for example, the strength of each externality is endogenous, and affected by the existence of the other two sources of externalities.

attained when the free-riding attempt was unsuccessful, and providing the public good results in an intermediate payoff. The net payoffs are constructed under the following benefit/ cost values parametrized by the free-parameter c_1:

- Base level of payoff for any player $= 10 + c_1$
- Public good benefits applicable above base payoff = 20. The benefits from the public good are accrued only as long as you, or someone you are directly connected to in the network, has provided the public good.
- Cost of linking to another player $= c_1$
- Cost of linking to $P0$ is c_2 where $c_2 = c_1 + (30 - B) > c_1$.
- In addition, players receive a bonus of $H \in \{0, 4\}$ points for every link made *to* their node.

Any value $c_1 \geq 0$ generates the same distribution of final payoffs for the ND game. The values of B and H are fixed before the game begins and are common knowledge. In the $H > 0$ treatments, we reintroduce Externality 2 but, importantly, maintain the exogeneity of Externality 1 and Externality 2.

All simultaneous network games with a candidate equilibrium in asymmetric strategies and symmetric players have multiple equilibria, as permuting the player identities of any equilibrium network creates another new candidate equilibrium. This confounds strategic uncertainty arising from the multiple rationalizable strategies (arising from the multiple equilibria) with other behavioral forces that might be present in the game. We therefore consider two timing structures for the ND: a simultaneous game, and a sequential game where it is common knowledge that P1 moves first, then P2, then P3, and finally P4 moves. Our sequential game has a unique sequentially rationalizable outcome (and thus a unique SPNE) and hence bypasses the confound of multiple equilibrium predictions.

In the remainder of this section, we identify equilibrium and rationalizable strategies in the ND under selfish and other-regarding preferences. In section 2.1 we characterize the Nash equilibrium of the ND under both self-interested and inequality averse preferences. We also characterize the set of extensive form rationalizable strategies. In section 2.2 we allow for agents to face uncertainty over the preferences of the other players in the game, imposing minimal assumptions on the priors that agents hold. In section 2.3 we discuss the implications of impure altruism for behavior in the ND. Readers who are interested only in our behavioral hypotheses, and not their derivation, may proceed directly to section 3.

2.1 Equilibrium and rationalizability

In this section we provide a characterization of expected behavior in the ND under two standard solution concepts, equilibrium and rationalizability. All proofs are relegated to the appendix. We consider three differing assumptions on preferences: self interested preferences, Fehr and Schmidt (1999) inequality aversion and Bolton and Ockenfels (2000) inequality aversion.

For Propositions 1 to 3 we assume standard self interested Expected Utility preferences. In any pure strategy Nash equilibrium of the ND exactly one player provides the public good and all other players link to the public good provider. Thus, every pure strategy Nash network is a star network.

Proposition 1. *The ND has four pure strategy Nash equilibria: for $i \in \{1, 2, 3, 4\}$, Pi links to P0 and every other player links to Pi. All outcomes are rationalizable.*

The intuition for the Nash equilibrium result is straightforward. When a player is providing the public good, the best response for everyone else is to free-ride off that person's public good provision. If no other player is providing the public good (and instead connecting to you), then the best response is to provide the public good. An immediate consequence is that all strategies and outcomes are rationalizable in the simultaneous ND.

In the sequential game, there are two distinct cases. When $B + 3H < 30$ the unique SPNE outcome is for the first three players to link to P4 and for P4 to link to P0; when $B + 3H > 30$ the star networks centered on either P1 or P2 are both consistent with SPNE play.

Proposition 2. *In the sequential ND the sub-game perfect Nash equilibrium outcomes are:*

- *a star network centered around P4 if $B + 3H < 30$.*
- *a star network centered around P1 or P2, if $B + 3H > 30$.*

For the case where $B + 3H < 30$ it is more valuable to be on the periphery of the star network than to be in the center of the network. In the sub-game perfect equilibrium the first three players connect to P4, leaving P4 with a unique best response of providing the public good for the entire group. For the case where $B + 3H = 36 > 30$ the center of the star is more lucrative than the periphery. The equilibrium where P1 provides the public good and all others link to P1 is intuitive. The equilibrium where P2 centers the star can be sustained by the (credible) threat that P2, P3 and P4 will form a mini-star around P2 if P1 deviates to providing the public good initially.

It is appropriate to use an extensive form variant of rationalizability to study our sequential game. To do so we use Extensive Form Rationalizability or EFR, as defined in proposition 2 of

Battigalli (1997)[11]. Given the simplicity of our game, where each player only moves once, we do not need nor reproduce the full machinery of Battigalli (1997). The procedure for finding the rationalizable set proceeds by iteratively restricting the set of beliefs that a player may hold about the behavior of other players. At each step, the player discards beliefs that are not consistent with sequentially rational behavior by the other players. As an example after the first iteration, EFR would restrict P1, P2 and P3 to believe that P4 will provide the public good with probability 1 whenever no previous players have provided the public good.

Proposition 3. *In the sequential ND the set of EFR outcomes are:*

- *a star network centered around P4 if* $B + 3H < 30$.
- *a star network centered around P1 or P2, or P1 and P2 both link to P0 and P3 and P4 link to either P1 or P2, if* $B + 3H > 30$.

The equilibrium in each of our treatments generates an asymmetric payoff profile across players. If subjects are sufficiently inequality averse then the star network may not be an equilibrium network structure: subjects may prefer the lower but more equitable payoffs that arise from a network when the public good is not provided. We show this in Propositions 4 and 5 using two popular models of other-regarding preferences, the Fehr and Schmidt (1999) model and the Bolton and Ockenfels (2000) model. Similar results would also hold for other models of other-regarding behavior including, for example, Charness and Rabin (2002).

In the Fehr and Schmidt (1999) model of inequality aversion, a subject's utility is a function of their own material payoff, x_i, and the material payoffs of the other players in their group, $\{x_j\}_{j \neq i}$.

$$U_i(x_i, \{x_j\}_{j \neq i}) = x_i - \sum_{j \neq i} \frac{\alpha \max\{x_j - x_i, 0\}}{3} - \sum_{j \neq i} \frac{\beta \max\{x_i - x_j, 0\}}{3}$$

where α is a parameter that measures the agent's disutility from disadvantageous inequality, and β is a parameter that measures the agent's disutility from advantageous inequality. It is standard to assume $\beta < 1$ and $\beta \leq \alpha$.

[11] Following Battigali (1997), we will deal with indifferences and mixed strategies being played by any player-type by representing them as a probabilistic conjecture held by others about the action played by that player-type. For example, if a particular Player 3 type is indifferent between two actions and can potentially mix beween those two strategies, then other Players would have beliefs on those two possible actions being played by that Player 3 type.

Bolton and Ockenfels (2000) provide a tractable motivation function for an individual as

$$U_i(x_i, \{x_j\}_{j\neq i}) \quad = \quad ax_i - \frac{b}{2}\left(\frac{x_i}{x_i + \{x_j\}_{j\neq i}} - \frac{1}{n}\right)^2$$

where n is the total number of individuals interacting and a player's type is characterized by a/b, the ratio of weights that are attributed to the pecuniary and relative components of the motivation function.

Proposition 4. *If all players are sufficiently inequality averse such that*

- $6\alpha + \frac{(2+H)\beta}{3} + \frac{2\beta}{3} \geq B - 10$ *in the Fehr-Schmidt model, or*
- $\frac{a}{b} \leq \frac{(\frac{B+H}{B+3H+50} - \frac{1}{4})^2}{(B+H-10)}$ *in the Bolton-Ockenfels model*

then the ring network is an equilibrium in the simultaneous ND.

In the sequential treatment, the behavior of P4 is of particular interest

Proposition 5. *In the continuation game of the sequential treatment where the first three players have connected to the fourth player,*
i) Under the Fehr-Schmidt model, P4 connects to one of the first three players if $B + 3H < 30$ *and she is sufficiently averse to self-harming inequality (and additionally does not mind self-serving inequality in the case of* $H > 0$*). Otherwise P4 connects to P0.*[12]
ii) Under the Bolton-Ockenfels model, P4 always connects to P0 for the treatments with $H = 4$*. She connects to one of the first three players if she is sufficiently concerned about fairness and if* $H = 0$*.*[13]

2.2 Rationalizability and uncertainty over preference-types (with inequality averse players)

There are three channels through which non-selfish preference types[14], for example inequality aversion, can generate outcomes different from those predicted by Nash equilibrium with self interested preferences: First, the subjects could themselves be inequality averse, second they could believe (correctly or falsely) that other players are inequality averse, and third, they could believe that

[12] If $B = 24,\ H = 4$ she must always link to the outside node, irrespective of her preferences in the Fehr-Schmidt class.

[13] These qualitative predictions hold even if one uses the most general form of Bolton-Ockenfels preferences.

[14] A "preference type" determines how the set of payoffs map to a player's final utility.

other players hold beliefs (including higher order beliefs) about the existence of inequality averse behavior. The uncertainty about preference types can be modeled via the use of the Harsanyi transformation: Nature's move determines the preference type of each player in a game of complete but imperfect information. The players are aware of their own preference type, but as is common in the rationalizability literature, they are allowed to disagree on first and higher beliefs they hold about the types of others (i.e., the beliefs about nature's move). For example, as a player in the game, my beliefs about the public good provision of player i could depend on "my beliefs about her beliefs about the public good provision of others", and, on "my beliefs about her beliefs about the beliefs of others about the public good provision of others", and, so on, and such higher order beliefs might be incorrect.

In the following, we study the set of rationalizable outcomes under behavioral uncertainty. The results developed here do not rely on the players themselves holding non-selfish preferences (see Assumption A1), neither do they require players believing that others have non-selfish preferences; it is sufficient that players (perhaps wrongly) allow for the possibility of non-selfish preferences in their higher order beliefs. For this reason we use the term *behavioral uncertainty* to imply both the Knightian unpredictability experienced by any player about other player actions conditional on preference types, and also the uncertainty about what the other players might believe about each others' realized type.

We assume that, in our sequential game, Nature moves first and decides the preference-type of Players $1, 2, 3, 4$ independently.[15] We impose the following restriction on preference-types:

Assumption on admissible preference-types A1 (feasible moves by Nature): All preference types strictly prefer any outcome where they receive 30 or more points over any outcome where they receive 24 or less points.

Assumption A1 places only weak restrictions on preferences. Clearly, self-regarding preferences are consistent with A1. The assumption is also consistent with the canonical Fehr-Schmidt model, subject to a parameter restriction $\beta \leq \frac{9}{23}$. That is, our assumption restricts the degree of aversion to advantageous inequality, but places no restriction on the degree of disadvantageous inequality aversion.

Players do not see Nature's move (other than the information about their own type), and they hold beliefs about the preference-types the other three players would have. We allow all such

[15] Note that we do repeat the application of rationalizability to our simultaneous game with inequality averse players: clearly, all strategies are rationalizable for at least some preference type.

possible beliefs as long as they satisfy the following full-support assumption.

Assumption on Beliefs B1: *All types of all players, assign some non-zero probability to the possibility that any other player would have an inequality averse preference type that prefers the outcome (10+3H,10,10,10) over the outcome (B+3H, 30, 30, 30).*[16] *Higher order beliefs also assign a non-zero probability to all lower order beliefs assigning a non-zero probability to this "averse to free-riding" type.*[17]

This means that as a player in the game, I must assign a non-zero probability that any other player is of this inequality averse type, and I must believe that others believe with non-zero probability that any player could be an inequality averse type, and I believe that others believe that others believe that others are an inequality averse type with non-zero probability, and so on. This assumption implies that agents will prefer strategies that deliver a good outcome with certainty over strategies that deliver the same good outcome only when others are fully self interested (and deliver a bad outcome otherwise).

Given all players move at a single information set for any history, their past actions have no bearing on future play conditional on reaching the current information set. Thus, any updating or forward induction (or lack thereof) on player-types based on past actions is inconsequential to our analysis and is hence omitted. This simplifies the analysis substantially relative to the general case of Extensive Form Rationalizability studied in Battigalli (1997). The only consequence of updating on types of players is that at any information set which is reached after Player i has already played, all players must have degenerate beliefs on the action taken by i.

Proposition 6. *In the sequential ND with $B + 3H \leq 24$, there exist extensive form rationalizable outcomes of the game and realizations of preference types, where the first player connects to one of P2, P3 or P4 and receives a payoff of* 10*. This holds even when the realized type of all the players has no other-regarding behavior, and assumption B1 holds.*

The intuition behind the constructive proof is that P1 believes that it is common knowledge that P4 will play the SPNE with a high probability, and so links to P4. Both P2 and P3 believe, and believe that each other believes, that P4 is inequality averse and will deviate from the SPNE with a high probability. Given this, P2 links to P3 and P3 links to P0. P4 then rationally links to P3, and P1 earns the minimum payoff of 10.

[16]The payoff profile (a, b, c, d) indicates that the player in question earns a points, and the other three players earn b, c and d points.

[17]The first order beliefs held by different players need not agree, and neither do the beliefs held by someone about different players be symmetric.

In our experimental implementation of the ND we do not observe nor control the beliefs that subjects hold about the preferences of others. Nor can we guarantee consistency of higher order beliefs across subjects. We consider behavioral uncertainty to be the natural condition of a laboratory experiment, and seek to understand what strategies subjects might implement to mitigate their exposure to behavioral uncertainty.

Proposition 7. *Assume A1 and B1. Then, in the sequential ND, P1 connecting to P0 is a rationalizable strategy. Further, if $B + 3H \leq 24$ then linking to P0 guarantees the first player a payoff of B+3H in any rationalizable outcome.*

Propositions 6 and 7 imply that, when considering rationalizable outcomes for P1, linking to P0 guarantees a "safe" intermediate payoff and linking to another player generates a range of possible payoff outcomes, including the worst possible outcome. A P1 who is aware of the behavioral uncertainty, and is pessimistic about the actions of later moving players is, therefore, other things equal, more likely to link outside than an uncertainty neutral P1. Thus, under a hypothesis that the perception of uncertainty is heterogenous across subjects, we should expect some P1 to consistently link to P0 and others to consistently link to another player. Similar intuition also holds for the simultaneous version of the ND, where linking to P0 guarantees a payoff in the range $[B, B + 3H]$ and linking to another player generates a payoff in the range $[10, 30 + 2H]$.

2.3 Impure altruism

Impure altruism has a long history in the study of public goods, with Andreoni (1989) introducing the notion that an agent may value her own contribution to a public good over and above the total public good contribution to the society.[18] We assume that, in our local public good environment, an agent that exhibits impure altruism will experience an extra "warm glow" from providing the public good to the other agents. We therefore assume that a subject with impurely altruistic preferences prefers, as their first best outcome, the outcome where they earn $B + 3H \leq 30$ and every other player earns 30, for large enough values of B and H. Otherwise, their preference ordering is identical to those of a fully self interested player. In this case the unique equilibrium in the sequential ND, under impure altruism and for large enough values of B and H, is for the first player to link outside and all other players to link to P1.

[18] More recently, Bergstrom et al. (2015), classify between one-sixth and one-third of subjects as "let-me-do-it" types in a volunteer's dilemma environment.

3 Experimental Design

In our experimental implementation of the Network Dilemma we refer to P0 as the blue node, and the subject nodes (P1 to P4) as red nodes, reflecting the color scheme used on the subject's terminals. The simplified action space, where the agent's only action is to form a link to one other node, makes the environment particularly easy for subjects to understand.

The payoff structure we employ, as explained to subjects, is simple:

- If a subject links to the blue node (P0) then the subject earns $B \in \{12, 24\}$ points.
- If a subject links to a red node, and that node links to the blue node, then the subject earns 30 points.
- If a subject links to a red node, and that node links to a red node, then the subject earns 10 points.
- For each node that links to a subject, the subject receives a bonus of $H \in \{0, 4\}$ points.

We ran 8 sessions with 16 subjects per session.[19] Within each session, subjects were divided into 4-player, 5 node network groups. Subjects played the Network Dilemma 100 times, divided into four blocks of 25 rounds each. Groups and player labels within groups (i.e. P1, P2, P3 and P4) were fixed within each block and randomly shuffled between blocks. We converted points to dollars at an exchange rate of 100 points equals $1. This implies that a subject who earns 24 points in every round will earn $24 (plus a $5 show up fee). We implemented 4 different parameter orders, as outlined in Table ??. The timing (sequential or simultaneous) and H are varied within session, while B is varied between sessions. Our design allows us to test for order effects of subjects playing the $H = 0$ or $H = 4$ treatment first. We ran the simultaneous blocks before the sequential blocks in every session to avoid the sequential treatment (where the SPNE may be focal) contaminating the symmetric simultaneous treatment.

[TABLE 1 ABOUT HERE]

[19] Sessions were conducted using oTree (Chen et al., 2016) and cytoscape (for displaying the network graphs to subjects) at Purdue's Vernon Smith Experimental Economics Laboratory (VSEEL) in November-December 2018 and January 2019.

3.1 Hypotheses

While summarizing our behavioral hypotheses, we outline the solution concept (equilibrium vs rationalizability) and assumptions on population preferences behind each hypothesis.

Hypothesis 1 (Modal networks under self-interested play): Star networks should be common in both Simultaneous and Sequential treatments, with the relative frequency of star networks weakly higher in the Sequential treatment.

Under self interested play, star networks are the only possible Nash equilibrium outcome in pure strategies in the Simultaneous Treatment, and are *the only rationalizable outcome in the Sequential Treatment* (for $B + 3H < 30$, see Proposition 3). In comparison, there are many rationalizable outcomes in the Simultaneous Treatments, which suggests that star networks should be weakly more common in the Sequential treatment.

If all subjects are sufficiently inequality averse, then the star network is not an equilibrium outcome anymore, instead the ring network (as shown in Proposition 4) forms a pure strategy equilibrium. But the mere existence of inequality averse players does not rule out the occurrence of star networks: In a group of four players, as long as there is at least one player who does not mind the self-harming inequality in centering a star network, and the three other players do not mind the self-serving inequality of being the periphery in a star network, there is at least one star network which is an equilibrium. Further, it is more likely that the ring network is an equilibrium of the stage game when B is low. This gives us our next hypothesis:

Hypothesis 1' (Modal networks under strong inequality aversion): The ring network should be common in both Simultaneous and Sequential treatments, and more frequently observed in the $B = 12$ treatments.

Now we state two more behavioral hypotheses that are related to self-interest vs inequality aversion.

Hypothesis 2 (Behavior of P4 under self-interest in sequential treatments): In the Sequential treatment if the first three players connect to P4, then P4 connects to P0.

Hypothesis 2' (Behavior of P4 under inequality aversion in sequential treatments): In the Sequential treatment, if if the first three players connect to P4, then P4 does not connect to P0. This effect may be stronger when the payoff inequalities are high (i.e, when $B + 3H$ is small) or/ and when $H = 0$.

Hypothesis 3 (Center of star under rationalizability and self-interest in sequential treatments): Under the assumption that subjects are purely self interested, we expect the star network to contain P4 at the center when $B+3H \leq 30$ and either P1 or P2 at the center otherwise.

Hypothesis 3' (Persistent public good providers under behavioral uncertainty or impure altruism): When acting under aversion to behavioral uncertainty or under impure altruism, subjects would link to the outside link if no one has connected outside previously in the sequential treatment. Thus, we expect the star network to center around P1.

This follows from our analysis and discussion in Subsections 2.2 and 2.3.

Hypothesis 3" (Persistent public good providers under behavioral uncertainty or impure altruism): Any subject working under aversion to behavioral uncertainty or under impure altruism, would persistently link to the blue node in the simultaneous treatment.

We discuss the results related to Hypothesis 3" in section B.2.

4 Results

The data exhibits no significant order effects, so we pool the data across all sessions for the following analysis. We begin by examining the types of networks formed in section 4.1, and then discuss the behavior of P4 in section 4.2. Section 4.3 outlines which equilibrium is selected in the sequential version of the game. Section 5 introduces a new diagnostic treatment to further separate our hypothesis.

4.1 Types of networks observed

Tables 2 and 3 present the most commonly observed networks in each of our treatments. In all treatments the star network is the most commonly observed network, with the proportion of star networks ranging from 19% in the simultaneous $B = 12, H = 0$ treatment to 80% in the sequential $B = 24, H = 0$ treatment. Star networks are more than twice as common in the sequential than simultaneous treatments, confirming hypothesis 1.

The ring network, which is predicted by hypothesis 1', is observed in only 14 out of 3200 rounds. If, however, we expand our focus to any network that is payoff equivalent to the ring network (i.e. include all six network structures where no one links to the source node in treatments with $H = 0$) then we observe networks with a payoff of vector of $(10, 10, 10, 10)$ a total of 85 times in the sequential $B = 12, H = 0$ treatment and 132 times in the simultaneous $B = 12, H = 0$ treatment. To determine whether these networks are an equilibrium phenomenon, or a result of miscoordination, we consider the stability of each network.[20]

A network is strictly stable if a group forms the exact same network (for example, star networks with the same central node) in consecutive rounds; we measure stability as the percentage of rounds $t + 1$ which have the same network as in round t for $1 \leq t \leq 24$. The star network is the only network that is strictly stable more than 50% of the time, and it exceeds that threshold in 7 of the 8 treatments. In the only treatment where the star network is stable less than half the time, the sequential $B = 12, H = 0$ treatment, the stability of the star network increases to 68% if we relax the definition of strict stability to allow for a relabeling of nodes/players. This is caused by groups who systematically rotate the center of the star in the sequential $B = 12, H = 0$ treatment. This turn taking behavior is not regularly observed in any of the other treatments.[21] An arrangement where the star network is formed every round, but the four players take turns in being the central node, indeed serves an inequality-aversion motive, as it ensures that payoff differences between players are minimized, while maintaining payoff efficiency.

Of the six network structures where no player links to the source node, none is strictly stable more than 8% of the time in any treatment, nor stable more than 14% of the time. The star network is played regularly as a stable network in all treatments, but the ring network (nor any of its payoff equivalent variants) is not played repeatedly in any treatment. The evidence favors hypothesis 1 over hypothesis 1', with hypothesis 1' receiving some support only in the $H = 0$ treatments.

[TABLE 2 ABOUT HERE]

[TABLE 3 ABOUT HERE]

4.2 P4 behavior in the sequential treatment

As an additional test for inequality aversion driving the comparative statics of efficient play across the B, H treatments, we analyze the behavior of the fourth player in the sequential treatment (table 4) when the first three players have all linked to P4. We see that in the $H = 0$, $B = 12$ treatment, more than half of the time fourth players in this situation don't link to the source node, thus

[20] If the same network is played repeatedly, then we conclude that that network is being played as a stationary equilibrium.

[21] We observed rotation in 5 out 7 groups who exhibited sustained cooperation in the $B = 12, H = 0$ sequential treatment. In the sequential $B = 12, H = 4$ and $B = 24, H = 0$ treatments rotation was observed in only 3 out of 29 groups who exhibited sustained cooperation.

rejecting an unequal but payoff maximizing outcome for a more equal outcome. Only 10% of the timefourth players exhibit the same behavior in the $H = 4$, $B = 12$ and $H = 0$, $B = 24$ treatments, where the equilibrium outcome is more payoff-equal. Subjects in the $B = 12, H = 0$ treatment are significantly less likely to link to the source node ($p = 0.001$ and $p = 0.010$ when compared to the $B = 12, H = 4$ and $B = 24, H = 0$ treatments, respectively, with standard errors clustered at the subject level). These results are consistent with inequality aversion being a significant driver of behavior *only in the* $H = 0$, $B = 12$ *treatment*, when the relative inequality is higher.[22]

This result provides evidence in favor of hypotheses 2' in the high inequality treatments, but not in the low inequality treatments. A comparison of P4 behavior across the $B = 12, H = 0$ and $B = 12, H = 4$ treatments suggests that the Bolton-Ockenfels model may better capture the inequality aversion of our subjects (compared with the Fehr-Schmidt model).

[TABLE 4 ABOUT HERE]

4.3 Equilibrium selection in the sequential treatment

[FIGURE 1 ABOUT HERE]

Figure 1 documents the relative frequencies of star networks and who the network was centered around, for the sequential treatments. Recall that the two treatments with $B = 12$, and the $B = 24, H = 0$ treatment have, under an assumption of self-interested preferences, a unique subgame perfect equilibrium in each round which involves the first three players linking to the fourth player and the fourth player linking to the source node, i.e, the star is centered around Player 4.[23] As seen in Figure 1, the rate of subgame perfect equilibrium play in the sequential treatments is low.The most frequently observed star, in these treatments, is a star centered on the first player.

[22]The discussion in this paragraph does not account for selection effects: P1 may be less likely to link to a P4 that has previously shown an unwillingess to cooperate. Accounting for selection effects, by taking the subject-wise mean of the proportion of rounds where P4 links to P0, does reduce the measured level of public good provision by P4 in the two $H = 0$ treatments (selection effects do not appear to play a role in the $H = 4, B = 12$ treatment). However, even after accounting for selection effects, P4 is willing to provide the public good more than half the time in the $B = 24, H = 0$ treatment, and nearly 40% of the time in the $B = 12, H = 0$ treatment. Selection effects on P4 behavior in the two treatments with $B + 3H = 24$ are muted by the large proportion of groups that never have all of P1, P2 and P3 link to P4 – these groups select away from linking to P4 before observing P4 behavior.

[23] It is therefore also the case that the unique SPNE for the 25 round supergame involves the first three players linking to the fourth player and the fourth player linking to the source node in each round.

The data cannot be explained by P1 testing, or exploring, the willingness of other players to center a star before reluctantly linking outside – the typical first player links to the source node from the very first round. Define a group to be cooperative if they form an efficient star in at least 5 consecutive rounds, and as stationary if the exact same network is formed in each of the 5 rounds. 26 of the 32 groups in the $B = 24, H = 0$ and $B = 12, H = 4$ treatments are stationary and cooperative, of which 17 cooperate on a P1-star, and a further 3 groups are cooperative but not stationary (they rotate the central member of the star).[24] Restricting attention to groups that were both stationary and cooperative, the modal number of rounds of experimentation for P1 before volunteering to center the star is 0. The median number of rounds is 1. That is, more than half of the P1 who center a cooperative star are willing to do so before establishing whether there is another player in the group who is willing to center the star (which would require at least three rounds of experimentation).

The most common network is a star network in all treatments, and the most frequent star network is a network around P1. This result provides evidence in favor of hypotheses 3', but not hypothesis 3. That is, behavior is not consistent with self-interested subgame perfect equilibrium, but modal behavior is consistent with a model of impure altruism or of sequential rationality with inequality aversion and behavioral uncertainty.

We know (see Proposition 3) that a star network around P4 is the unique rationalizable and subgame perfect outcome, for $B + 3H < 30$, when all players are self-interested. Given that, on aggregate, P4 links to the outside node often enough to justify subgame perfect behavior from P1, P2 and P3 (see table 4) it is interesting to consider which players deviate from subgame perfect behavior and whether such deviations are empirically justified.

[TABLE 5 ABOUT HERE]

Disagreement of beliefs: Under the behavioral uncertainty hypothesis, the final column in Table 5 reveals that players often disagreed on their beliefs about the motivations and actions of later moving players. For example, in the $B = 12, H = 4$ and $B = 12, H = 0$ treatments, for one third of the occasions where P1 thought it was worth connecting to P4, P2 did not do the same. Had both P1 and P2 held the same first order beliefs about P3 and P4's actions (and motivations), such a disagreement of actions would not have been possible.[25] Similarly, in the $B = 12, H = 0$

[24]There is one group that initially cooperates on a P2-star for periods of 6 and then 11 rounds, before switching to cooperating on a P1-star for the final 5 rounds. All other stationary cooperative groups only cooperate on a single network configuration.

[25]Given that subjects are randomly assigned to roles, the observed behavioral difference should not be caused by

treatment, in one third of the occasions where P1 and P2 thought it was worth connecting to P4, P3 did not do the same.

Importance of higher order beliefs: Table 5 shows the empirical payoffs earned by each player while playing the SPNE dictated strategy vs deviation strategies, conditional on all previous movers playing the SPNE dictated strategy. It is clear that deviations from the SPNE by player 1 are typically not costly in any of the three payoff configurations. Given the empirical proportion of deviation from the SPNE path by the following two players, P1 should expect play to reach P4 along the SPNE path in (100-13)%×(100-13)%=76% of rounds in ($B = 24, H = 0$) and as little as 46% of rounds in ($B = 12, H = 0$). *Thus, comparing with Table 4, one can see that the main source of uncertainty for P1 while contemplating playing her SPNE strategy, is not whether P4 will play their part in the SPNE, but whether P2 and P3 will play their part, especially in the treatments with B+3H=24.* Or, alternatively, P1 might not be concerned that P4 is inequality averse, but instead might be concerned about higher order beliefs: does P3 believe that P4 is inequality averse, and does P2 believe that P4 is inequality averse or believe that P3 believes that P4 is inequality averse?

The deviations from SPNE behavior by P2 and P3 are, empirically costly for them on average.[26] We can therefore attribute the failure of SPNE, at least in part, to P3 and P2 holding incorrect beliefs about the behavior of P4. Thus, despite P1 being prima facie the main cause of deviations from SPNE play, a deeper analysis suggests that the behavior of P1 can be justified by holding (correct) beliefs about off equilibrium behavior of P2 and P3.

The preceding discussion also highlights the central tension between the strategic advantage of early movers (being able to commit to providing the public good or not) and the informational advantage of later players (faces no uncertainty over other's actions). The substantial rate of deviations from SPNE behavior by P2 and P3 documented above suggest that the informational advantage of later players may outweigh the strategic advantage of the first mover. Table 6 documents that this is, indeed, the case for the three treatments with $B + 3H < 30$. However, for the case of $B + 3H = 36$ the strategic advantage of P1 dominates: by providing the public good immediately, P1 is able to induce the other players to form a star around P1.

[TABLE 6 ABOUT HERE]

differences in preferences across roles.

[26] Learning does not appear to play a large role here. The only appreciable difference from restricting attention to the final 10 rounds is that P2 and P3 play the SPNE substantially more often in the $B = 24, H = 0$ treatment, which improves the prospects of P1 if they link to P4. P1 does not respond to this change in behavior, however, and actually plays their SPNE strategy less often in the final 10 rounds.

5 A diagnostic treatment

To separate the channels of *Impure Altruism* and *Behavioral uncertainty*, one needs a treatment that would remove either of the subject motivations. One way to do this would be to make P1 (who are highly likely to volunteer in our base treatment results) play against the distribution of actions taken by P2, P3 and P4 from past sessions, thus removing the altruistic possibility that an earlier moving player can "help" the pecuniary rewards of a later moving player through her actions. But, this would still not be a clean test, as it is possible that volunteering is an act of virtue-signaling to self, and such motives might still remain undiminished under such an alternative treatment.

Hence, we take a different approach: To test whether the behavior of volunteers (for example, P1's behavior in the sequential treatment) is being driven by aversion to uncertainty or impure altruism we designed a new treatment that substantially lowers the degree of behavioral uncertainty experienced by the first mover. This *forced linking* treatment differs from our *baseline* treatment in only one dimension: if any set of three players all link to the remaining player, then the remaining player's action in that round is overridden and the she is forced to link to the blue node. We ran two sessions of the *forced linking* treatment using four blocks of 25 rounds each, with two blocks of simultaneous games followed by two blocks of sequential games, reflecting the order of the *baseline* treatment. The *forced linking* treatment used only two sets of parameters: $B = 12, H = 4$ and $B = 24, H = 0$; the two parameter sets where we observed the most non-subgame perfect equilibrium behavior.

5.1 Behavioral predictions

The *forced linking* treatment does not change the expected behavior under an assumption of impure altruism: the payoffs are exactly the same as the *baseline* treatment so that any player who prefers to provide the public good in the *baseline* treatment also prefers to provide the public good in the *forced linking* treatment. Therefore, the requirement that a subject must provide the public good when all others are linked to the subject does not restrict the best response behavior of the subject.

However, for our model of rationalizability with inequality averse players who face behavioral uncertainty, the *forced linking* treatment changes the predicted behavior substantially, and ensures favorable outcomes for P1 in all subgames following P1 connecting to another player. For example, P1 connecting to P4 results in one of two possible (sequentially rationalizable) outcomes in the ensuing subgame: either P2 foresees that her connecting to P4 would induce P3 to follow suit thus ensuring the first three players the optimal outcome from the game (even under other-regarding preferences, see assumption A1). Alternatively, P2 could connect back to P1 foreseeing that thereafter P3 and P4 would follow suit (as it would be the only remaining way of achieving P3 and P4's favorite outcome under A1), and then the rules of the *forced linking* game would enforce a star network centered around P1 with P1's action being revised to link to P0. Under sequential rationalizability, in the *forced linking* treatment Player 1, after linking to another player, can expect the gameplay to follow a path where she is either the center or periphery of an efficient star, and gets no less than she would receive from linking to P0 directly, despite the uncertainty (first and higher order) about the preference types of players P2, P3 and P4 allowable under A1. In other words, after restricting to rationalizable strategies, P1 connecting to another player weakly dominates P1 connecting to P0. We show this formally in the following two propositions.

Proposition 8. *Assume that preferences satisfy assumption A1, and beliefs satisfy assumption B1. Then, in the* forced linking *treatment, all sequentially rationalizable outcomes are efficient star networks. The first player ensures herself a payment of at least* $B + 3H$*, and at most* 30 *in any sequentially rationalizable outcome of the game.*

Recall that in the context of the linking game, an outcome is the set of all payments received by the four players in the game. In the following proposition we use weak dominance to compare outcomes across treatements and strategies. Clearly this is a particular *partial* order over sets of outcomes, and we will use this in Proposition 9 to suggest why certain actions are more likely to be seen both within and across treatments, under behavioral uncertainty. We will need one more weak assumption on preferences for Proposition 9.

Assumption on admissible preference-types A2 (Admissble strategies for Nature): All preference-types would prefer being a member of a star network (either at the center or on the periphery) strictly more than the allocation $(10, B + 2H, 30, 30)$.[27]

We use the notation O_j^i to denote the set of rationalizable outcomes for treatment $i \in \{\boldsymbol{s}tandard, \boldsymbol{d}iagnostic\}$ when the linking action taken by player 1 is $j \in \{\boldsymbol{b}lue, \boldsymbol{r}ed\}$.

Proposition 9. *Under preference assumptions A1 and A2, for Player 1, and with* $B + 3H \leq 24$
(i) The set of rationalizable outcomes O_r^d *in which P1 links to another player in the* forced linking

[27]This assumption places no further restriction on Fehr-Schmidt preferences beyond those imposed by assumption A1.

treatment weakly dominates the set O_b^d of rationalizable outcomes in which P1 links to P0.

ii) The set O_r^s of possible outcomes sequentially rationalizable by linking to another player in the baseline treatment does not dominate the set O_b^s of outcomes rationalizable by linking to P0, and the latter does not dominate the former either.

The *forced linking* treatment also introduces new SPNE relative to the *baseline* treatment: a star formed around any player can now be supported as an SPNE. This reduces the salience or focality of the star network centered around the fourth player, as it is no longer the unique SPNE of the game. Therefore, our new treatment forms a particularly stark test of the behavioral uncertainty hypothesis, as the equilibrium predicted by the hypothesis is no longer focal. Summarizing, if impure altruism is driving behavior in the sequential treatment, then we should expect to see the first player linking to the source node in the *forced linking* treatment as frequently as before. If, on the other hand, behavioral uncertainty is driving behavior, then we expect to see a star centered around the last moving player more often.

[TABLE 7 ABOUT HERE]

5.2 Results

Figure 2 shows the key result from the *forced linking* treatment: the modal outcome in the treatment with lower behavioral uncertainty is a star centered around P4. We reject the hypothesis that impure altruism was driving behavior in the *baseline* treatment – the first player does not want to be in the center of the star network in the new treatment. The reduction in behavioral uncertainty generates a huge shift in P1 behavior, as they move from linking to the blue node in the *baseline* treatment to linking to P4 in the *forced linking* treatment. Table 8 confirms that this large change in behavior is statistically significant: There are significantly more P4 stars, and significantly fewer P1 stars and non-equilibrium networks in the *forced linking* treatment.

[FIGURE 2 ABOUT HERE]

[TABLE 8 ABOUT HERE]

We also ran the *forced linking* treatment for the simultaneous treatment and summarize the results in the appendix.

6 Relationship to the prior literature

Choi et al. (2016) provides a thorough review of network experiments, with a focus on public goods in section 17.2.2.[28] Here, we restrict attention to those papers which are particularly related to our work. Most network experiments can be placed into one of two categories: experiments where the only subject decisions are related to the formation of links (Falk and Kosfeld (2012); Goeree et al. (2009) for example), and experiments where subjects play a game on a fixed network (Rosenkranz and Weitzel (2012); Charness et al. (2014); Mukherjee (2019) for example). A recent strand of public good experiments, discussed below, allow for simultaneous network formation and public good formation decisions.

Our experimental environment is related to the theoretical model of Galeotti and Goyal (2010) (henceforth GG), that explains how a majority of individuals might obtain the majority of their information from a small subset of the population ("Law of the few"). In their model, each agent must simultaneously decide how much information to provide and which links to form, and all connected individuals gain from the information provided. Thus GG can be interpreted as a model of investment in public good provision coupled with endogenous network formation. In our experiment, each agent decides whether to provide a binary public good (linking to P0) or which other link to form. This simplifies the action space by having the linking decisions also account for public good investment in a succinct way. Our environment is not a special case of GG: we assume a one-way flow of benefits along edges (GG assume two-way flow), and we decouple the cost of linking from the benefits of network centrality (GG models centrality benefits as a transfer from the source to termination node). One major point of comparison between the two frameworks is that in our environment the strict Nash equilibrium is also efficient, whereas, in GG, any equilibrium network with linking has an inefficient level of information acquisition. GG is therefore more suited to modelling the under-, or over-, provision of public goods than our environment. In an online appendix we provide a generalization of the ND that provides a natural point of comparison between the similarities and dissimilarities between the current framework and the Galeotti and Goyal (2010) model.

There are three experiments that have tested the GG model in the laboratory. Rong and Houser (2015) (henceforth RH) is the closest to our experiment, as they restrict public good provision to be binary. After normalizing payoffs the *Sim_B* treatment in RH is payoff equivalent to setting $B = 26$ and $H = 0$ in our experiment. The key difference is that subjects in RH may form many links and

[28]While thorough, more recent works such as Mukherjee (2019) are not included.

provide the public good at the same time, and are not restricted to forming only one link as they are in our experiment. In their *Sim_L* treatment they restrict each player to only one action (either forming one link or investing in the public good) but also impose a restriction that only a maximum of one player may invest in the public good – this design feature greatly enhances network efficiency. RH also implement a sequential version of their experiment, although the implementation (allowing one subject to change their action each round, while holding the remaining subject actions constant) is sufficiently different from our sequential treatment that comparison is difficult.

van Leeuwen et al. (Forthcoming) (henceforth vLOS study a GG model with discrete public good provision and incorporate status rents. The status rents in vLOS are sufficient to ensure that the central player earns more than the periphery, and the primary focus is on the relationship between status rents, the level of public good provision, and network structure. The status rents available to the central node induces competition via an over provision of public goods. vLOS also find that inequality aversion is relatively unimportant in a local public goods game, instead attributing the over provision of public goods to the effects of status competition.

Goyal et al. (2017) also study the GG model in the laboratory, and is the most faithful experimental implementation of the GG model to date. They vary the cost of linking and the heterogeneity of the cost of public good provision between treatments. Each round of a public good network game is played as a continuous game with a random stopping time, and only actions at the random stopping game are payoff relevant. This framing of the experiment may help to increase coordination, as the early seconds of each round can be used as a communication device.

Charness et al. (2014) (henceforth CFMS) studies the effects of network uncertainty on subject behavior. The key distinction between CFMS and the current paper is the source of uncertainty – in CFMS the uncertainty is exogenous, while here the uncertainty is endogenously generated from the behavior of other players. In both papers, however, uncertainty has the counter-intuitive effect of supporting, rather than hindering, equilibrium behavior. Carpenter et al. (2012) studies public good provision in incomplete networks where subjects can only monitor and punish connected members. Choi et al. (2019) studies network formation and specialization in large networks, and find evidence for "law of the few". Gachter et al. (2010) and Andreoni et al. (2002) study and compare simultaneous and sequential two-player public goods games with quasilinear utility. Both papers find substantial amounts of non-equilibrium behavior in their sequential treatments and attribute the non-equilibrium behavior to social preferences.

Relative to the previous literature, our design simplifies the decision making environment faced by subjects. In particular, we embed a binary public goods game into a pure network formation game – subjects only need to form links, rather than simultaneously form links and provide public goods. We limit our network effects to only immediate neighbors, and we restrict the flow of public good benefits to be unidirectional. These restrictions make the game easier to understand for our subjects, and simultaneously enable the use of revealed preference methodologies to infer the underlying preferences of our agents. Theoretically, as discussed in Section 2, our simplified design is equivalent to restricting the nature of the network externalities that can operate in our game.

7 Discussion and conclusion

Our paper, which embeds a public goods game into a pure network formation game, provides a link between the literature on pure network formation games and the literature on public good network games. We focus on understanding network-mediated public good externalities, and abstract away from other externalities in our experimental design. By using a sequential network formation treatment besides our simultaneous treatment, we are able to diminish the multiplicity of rationalizable strategies/equilibria, and thus are able to identify the effects of strategic uncertainty and other behavioral forces. While the previous literature on pure network formation finds a very strong effect of payoff inequality on behavior, we instead find that payoff uncertainty (from higher order beliefs) is a key determinant of behavior.[29] The public good payoff structure of our network game, which provides a "safe" option to those who volunteer to provide the public good helps to facilitate the formation of efficient networks.

In our baseline sequential treatment we find that, unless payoff inequality is extreme ($B = 12, H = 0$ treatment), subjects in the role of P4 (i.e. the last to move) are willing to accept a degree of payoff inequality in achieving the efficient equilibrium outcome. In the same treatment we observe subjects in the role of P1 (i.e. the first to move) volunteering to provide the public good and placing themselves in the center of an efficient star network. Turning to the baseline simultaneous treatment we observe the same willingness to volunteer helps facilitate the formation of star networks: repeatedly providing the public good allows an efficient star to form around the public good provider. Finally, in the forced linking treatment we identify that it is an aversion to uncertainty regarding the behavior of others, and not impure altruism, that is the source of the willingness of voluntary provision by early-moving players in networks of purely Public good externalities.

[29] Although not a pure network formation game, van Leeuwen et al. (Forthcoming) find in their network public goods game that a competition for status, rather than payoff inequality, is the key determinant of behavior.

Tying it back to the network mediated inter-country innovation-sharing example, our results suggest that

- The possibility of more equitable terms of trade (like in the $H = 4$ treatment, as compared to $H = 0$) in the face of innovation-sharing might avoid inefficient stand-offs.
- Similar to the repeatedly volunteering subjects at the start of repeated-interaction blocks in our simultaneous treatment, if newly formed governments (after a recent election or regime change) can credibly and persistently signal their type and intentions, that might help resolve an impasse over what roles their countries could play in sharing of technology and expertise.
- Simple interventions by international bodies like the WTO (like the one in the diagnostic treatment) might have a disproportionately large impact in avoiding uncertainty and inefficient stand-offs, and determining who plays what role (innovator vs technology receiver) in sharing of technology and expertise.

While the first is an insight we already know from a large varity of games (like the ultimatum game), we believe that the second and the third insights are novel.

A Proofs

Proof. [Proposition 1]

That the star network is a NE of the network formation game is self evident. To establish that there are no other pure strategy equilibrium, consider two cases. In the first, suppose that no player has linked to P0. In this case, any player who deviates to linking to P0 will increase their payoff by $B - 10 > 0$. In the second case, suppose that multiple players have linked to P0. At least one of the players will have a profitable deviation linking to the other player who linked to P0.
Given that there exists some equilibrium in which each player links to P0 or to any other player, it follows immediately that all strategies are rationalizable. □

Proof. [Proposition 2]

Note that in all subgames where no previous player has linked to P0 the best response for P4 is to link to P0. When $B + 3H < 30$ the star networks centered around P1, P2, and P3 are not subgame perfect. In each case, the center of the star has a profitable deviation of linking to P4, knowing that once P4 moves they will link to P0. The star network centered around P4 is clearly subgame perfect.

When $B + 3H > 30$ the center of the star earns more than the periphery. If a previous player has linked to P0 then the best response for P4 is to link a player who is linked to P0. The same is true for P3. Consider the subgames that begin with the move of P3. If only P1 has linked to P0, then both P3 and P4 link to P1. If only P2 has linked to P0, then both P3 and P4 link to P2. If both P1 and P2 have linked to P0, then P3 and P4 may link to either P1 or P2. Therefore, a star around P1 is a SPNE and is supported by P3 and P4 linking to P1 in the subgame where both P1 and P2 link to P0. A star around P2 is also a SPNE, and is supported by P3 and P4 linking to P2 in the subgame where both P1 and P2 link to P0. □

Proof. [Proposition 3] We will show the case for $B + 3H < 30$. The other case follows similarly. We start by restricting the beliefs that Players 1-3 can hold on the actions of a rational Player 4 in each potential information set where she moves, and this would also put restrictions on higher order beliefs.

Player 4:

After any information set where no one yet has connected outside: She connects outside as $B > 10$. At any information set where at least one player has previously connected outside: She would connect to an outside connecting player.

Next we restrict the beliefs that Players 1-2 can hold on the actions of a rational Player 3 after each possible history.

Player 3:

At any information set where no one has connected outside yet: Player 3 should connect to Player 4 (follows from our analysis of the beliefs Player 3 holds about Player 4's action in the ensuing game, which allocates 30 to Player 3 eventually), as connecting outside herself would get her no more than $B + H < 30$.

At any information set where at least one player has previously connected outside: She would connect to an outside connecting player, also because, connecting outside herself would get her no more than $B + H < 30$.

Next we restrict the beliefs that Player 1 can hold on the actions of a rational Player 2 after any possible history.

Player 2:

At any information set where no one has connected outside yet: Player 3 should connect to Player 4 (follows from our analysis of the beliefs Player 2 holds about Player 3 and 4's action in the ensuing game).

If the first player has previously connected outside, Player 2 would connect to her, because, connecting outside herself would get her no more than $B + 2H < 30$.

Player 1:

Player 1 would connect to Player 4 (expecting a payoff of 30 given her beliefs) because, connecting outside herself would get her no more than $B + 3H < 30$, and connecting to any other player gets her 10. □

Proof. [Proposition 4] Consider the ring network, where the following links are made: $1 \rightarrow 2$,$2 \rightarrow 3$,$3 \rightarrow 4$,$4 \rightarrow 1$. In the deviation case of connecting outside, the payoffs to the four players are $\{B + H, 30 + H, 10 + H, 10\}$, so that the differences in payoffs are $-(30 - B)$, $2 + H$, and 2. Now, no player wants to connect outside, given no one else has connected outside, as long as,

$$\begin{aligned} B + H - \frac{18\alpha}{3} - \frac{(2+H)\beta}{3} - \frac{2\beta}{3} &\leq 10 + H - 0 \\ \iff 6\alpha + \frac{(2+H)\beta}{3} + \frac{2\beta}{3} &\geq B - 10 \end{aligned} \tag{A.1}$$

Equation A.1 gives the necessary conditions[30], while $\alpha \geq \frac{B-10}{6}, \beta = 0$ is a sufficient condition. A player would also not be willing to deviate to link to another player, as that would weakly increase payoff inequality without improving the material payoff of the deviating player.

Similarly by the Bolton-Ockenfels model, no player wants to connect outside, as long as,

$$\begin{aligned} a(B + H) - b(\frac{B + H}{B + 3H + 50} - \frac{1}{4})^2 &\leq (10 + H)a - 0 \\ \iff b(\frac{B + H}{B + 3H + 50} - \frac{1}{4})^2 &\geq a(B - 10) \\ \iff \frac{a}{b} &\leq \frac{(\frac{B+H}{B+3H+50} - \frac{1}{4})^2}{(B - 10)} \end{aligned}$$

Now, $\frac{B+H}{B+3H+50} \notin (\frac{12+4}{12+12+50}, \frac{24}{24+50}) = (\frac{16}{74}, \frac{24}{74})$ and hence for $\frac{a}{b} \leq \frac{(\frac{B+H}{B+3H+50} - \frac{1}{4})^2}{(B+H-10)}$ the ring network would be an equilibrium. □

Proof. [Lemma 5] **Fehr-Schmidt Preferences:** Under $B + 3H < 30$, the fourth player does not link outside, as long as,

[30] The higher is α or β the more likely is that the necessary condition holds.

$$\begin{aligned} B + 3H - (30 - B - 3H)\alpha &\leq 10 + 3H - \frac{8\beta}{3}H \\ \Longleftrightarrow (30 - B - 3H)\alpha - \frac{8\beta}{3}H &\geq B - 10 \end{aligned} \tag{A.2}$$

Equation A.2 gives the necessary conditions[31], while $\alpha \geq \frac{(B-10)}{(30-B-3H)}, \beta = 0$ is a sufficient condition.

Under $B = 24,\ H = 4$, the fourth player always links outside, as linking inside reduces her pecuniary payments, as well as increases inequality.

Bolton-Ockenfels model: For the fourth player, here is how her payment as a proportion of total payments change in the different treatments:

[TABLE A1 ABOUT HERE]

Connecting intside can bring the proportional payment of P4 closer to $\frac{1}{4}$ only when $H = 0$. Thus, irrespective of the ratio $\frac{a}{b}$, P4 always connects outside when $H = 4$. When $H = 0$, $P4$ would connect inside when the ratio $\frac{a}{b}$ is small enough, i.e, when she cares enough about fairness.

□

Proof. [Proposition 6] □

We provide an outline of the existence-proof for $B = 24, H = 0$ using the admissible preference types, for simplicity. Similar examples can also be constructed for the other B, H treatment parameters used in the paper, in the range of $B + 3H \leq 24$.

Suppose all the players assign a small probability (equal to $p = \frac{1}{10}$ in case of beliefs held by 1 and $p = \frac{1}{3}$ in case of Player $2, 3, 4$) to the possibility of every other player is a preference type who would prefer (and hence choose) the $(10, 10, 10, 10)$ outcome to the outcome where she is the center of an efficient star, and with the complementary probability believe that all the players are self-regarding (i.e, take actions that maximize personal payoffs). Let P2 hold the correct belief about what P3 thinks about P4, but let P1 believe that P2 and P3 share P1's first-order beliefs about P4, and let P1 believe that P2 believes that P1, P2 and P3 hold the same first order belief on P4. Let the unspecified higher order beliefs of every player be consistent with their lower order beliefs. For simplicity, let the realized type of all the three players be that they only care about maximizing personal payoffs. There exists a rationalizable outcome where

[31] The higher is α or the lower is β the more likely is that the necessary condition holds.

(1) P1 connects to P4 because P1 wrongly believes that, (1.i) P4 will connect to P0 with probability 0.9 as long as all players connect to her, (1.ii) P3 believes <*1.i*> too, and hence connects to P4 as long as all previous players connect to P4, and (1.iii) P2 believes in both <*1.i*> and <*1.ii*> and hence also connects to P4 as long as P1 connects to P4.
(2) P2 correctly guesses that P3 would connect to P0 as long as neither P1 or P2 have done so (because she correcty guesses that P3 believes that P4 will connect to P0 with a probability of only $\frac{2}{3}$ after P1-3 connect to her), and hence P2 connects to P3 after P1 connects to P4.
(3) P3 connects to the outside node as long as neither P1 or P2 have done so.
(4) P4 connects to P3 if P3 is the only previous player who has connected to the outside node.
In this outcome, P1 earns 10. Note that this inefficient outcome occurred even though all players were self-interested and rational: the inconsistency of higher order beliefs across players is enough to prevent the formation of an efficient star network.

.

Proof. [Lemma 7]As long as the first player believes that there is a high chance of all the following players being averse to centering a star, she would find it worth connecting outside herself. Thus connecting outside is a rationalizable strategy.

Assume that P1 has connected to P0. Consider P4, who will always be able to guarantee a payoff of at least 30 by connecting to a node that is connected to P0. P3 knows this, and knows that P4's preferences satisfies Assumption A1. Therefore, P3 believes that P4 will only link to P3 if P3 links to P0, and that P4 will never link directly to P0. Therefore the best response of P3 is to link to a player who is already linked to P0, which guarantees P3 a payoff of at least 30.
P2 must also believe that P3 and P4 will only link to P2 if P2 links directly to P0, and that P3 and P4 will never link directly to P0. From assumption A1, P2's unique best response is to link to P1. Given this, the unique best response of P3 is to link to P1, and the unique best response of P4 is to link to P1. Conditional on P1 linking to P0, the unique rationalizable outcome is a star formed around P1 and P1 earns $B + 3H$. □

Proof. [Proposition 8] We start by restricting the beliefs that Players 1-3 can hold on the actions of a rational Player 4 in each potential information set where she moves, and this would also put restrictions on higher order beliefs.

Player 4:

After everyone has connected to her previously: She is indifferent between all actions as her action is overridden to connect outside.

At any information set where no one yet has connected outside: All Player 4 types would connect to the player who has already been connected-to by 2 previous players. If this is not available then she connects to the least connected player (for a high enough inequality averse player), or connect outside herself, or any other action, depending on her preference-type.

At any information set where at least one player has previoulsy connected outside: She would connect to an outside connecting player.

Next we restrict the beliefs that Players 1-2 can hold on the actions of a rational Player 3 after each possible history.

Player 3:

If both the first two players have connected to the fourth player: All Player 3 types should connect to the fourth player.

At any information set where no one has connected outside yet: If the first player has connected to Player 2, and Player 2 has not connected back to Player 1, then all types of Player 3 would connect to Player 2. If the first player has connected to Player $i \neq 2$, and Player 2 has connected to Player 1, then all types of Player 3 would connect to Player 1. If the first two players have connected to Player 3, then, all types of Player 3 are indifferent between all possible moves. If both player 1 and 2 have connected to each other, then Player 3 should be indifferent between connecting to one of those two players. In all the described scenarios, we are using the rationality of Player 3, and that she beliefs in the rationality of Player 4. In all other cases (and it will turn out that these never happen in rationalizable outcomes of the game), she connects to the last moving player, or the least connected player, or connects outside herself (depending on how optimistic her beliefs are on Player 4 connecting outside in the relevant information set and her preference-type).

At any information set where a player has previously connected outside: She would connect to an outside connecting player.

Next we restrict the beliefs that Player 1 can hold on the actions of a rational Player 2 after any possible history.

Player 2:

If the first player has connected to the third or fourth player, Player 2 is indifferent between connecting to that connected player, or connecting back to Player 1, and prefers those two actions to all other actions.[32]

[32] She does not connect to the the third or fourth (in case 1 connected to 4 or 3 respectively) due to our assumption on beliefs.

If the first player has connected to Player 2, then all types of Player 2 who put a positive probability on Player 3 following suit in case Player 2 connects to Player 1, would connect back to Player 1[33]. Otherwise, she would be indifferent between that and any other action.
If the first player has previously connected outside, Player 2 would connect to her.

Player 1:

Finally, considering the first moving player: Connecting to any player moving after her guarantees her at least $B + 3H$ in any rationalizable outcome, and as long as she *does not* put a degenerate belief on Player 2 connecting back to her at the ensuing information set, she should get no less than 30, and hence, strictly prefer to not connect outside. If she does believe that Player 2 will connect to herself, and Player 3 and 4 will follow suit, then she is indifferent between connecting outside and connecting to any other player. □

Proof. [Proposition 9] For (i), note that the set of rationalizable outcomes when P1 links to another player are a star around P1 or a star around the other player. When P1 links to P0, the only rationalizable outcome is a star around P1. The result then follows immediately from assumption A1.
For (ii), note that $O^{stand}_{another}$ includes the outcome and payoffs in the example constructed in the paragraph following Proposition 6, and also the outcome where Player 1 is the periphery of a star network involving all the players. Thus, $O^{stand}_{another}$ includes an outcome that is strictly better than all outcomes in $O^{stand}_{outside}$, and another outcome that is strictly worse than all outcomes in $O^{stand}_{outside}$. □

B Supplementary data appendix

B.1 Aggregate efficiency measures

We measure the degree of coordination in a network using a coordination index. The coordination index is more informative than the coarser measure of the proportion of star networks formed, as it is able to distinguish between networks that are closer to, or further from, the star network.[34] We construct our index using the following steps:

[33] See the assumption about representing mixed strategies as probablistic conjectures and the discussion of Player 3's move in the relevant information set.

[34] Our coordination index is a function of the network formed, and not the payoffs earned, to avoid generating a spurious correlation between the payoff parameters and our measure of coordination.

1. For each subject in each round, we assign a score of
   - 0 if they connect to the blue node,
   - 1 if they connect to a red node that connects to the blue node (a success),
   - -1 if they connect to a red node that connects to a red node (a failure).
2. For each group (that plays 25 rounds together under a particular BH-treatment), we average the score across all subjects in the group over the 25 rounds.

Our coordination index assigns a score of 3 (maximum score possible on our index) to a group that forms an efficient star in all 25 rounds, and a score of -4 (minimum score possible on our index) to a group that never has any subject link to the blue node. A group that has en equal number of successes as failures across the 25 rounds will earn a score of 0.

[FIGURE A1 ABOUT HERE]

Figure A1 displays box-and-whisker plots of the coordination index by treatment. The sequential treatments have higher index scores than the simultaneous treatments. In Column 1 of Table A2 we regress our coordination index on H, B and the timing of the game. We treat both H and B as categorical variables and use the $H = 0, B = 12$ simultaneous treatment as the omitted category. Standard errors are clustered at the group level. Table A2 confirms that all three of our treatment variables have a strong effect on the coordination index.

[TABLE A2 ABOUT HERE]

Our coordination index is increasing in both H and B, with the coefficient on B in the first column of table A2 larger than the coefficient on H, which is consistent with the hypothesis. The coordination index is also significantly higher in the sequential treatment, also consistent with the hypothesis.

It may appear surprising that coordination and cooperation rates are so high in the sequential $B = 12, H = 0$ treatment. Some groups in the sequential $B = 12, H = 0$ treatment were able to reduce aggregate payoff inequality by rotating the center of the star across rounds. Of the 7 groups in that treatment who were able to sustain the efficient star network for at least three consecutive rounds, 5 groups rotated the center of the star. This type of inequality avoidance cooperation was much less common in the other sequential treatments, and only observed once for a single three-round stretch in one of 64 groups in the simultaneous treatment.[35]

[35] 3 out of 15 cooperative groups exhibited center player rotation in the sequential $B = 12, H = 4$ treatment, and 2 out of 16 cooperative groups did so in the sequential $B = 24, H = 0$ treatment.

B.2 Persistence and willingness of public good providers

Next, we consider behavior in the simultaneous treatment in rounds prior to the *first* star network being formed. We seek to understand how cooperation might arise and identify who might end up in the center of a star network. It is plausible that subjects might play a mixed strategy, or independently play some rationalizable strategy every round, until a star network is formed, and keep playing the same action there onwards (thus making the star networks stable). Alternatively, some subjects may try and foster coordination by repeatedly linking to P0 and signaling that they are willing to center a star. Otherwise, some subjects may repeatedly link outside simply because they prefer the surety of receiving a payoff of at least B points. In each of the alternative cases we refer to these persistent public good providers as *early adopters*, even though the motivations of the early adopters differs in each case.

Let $b_{i,t}$ be an indicator variable that equals one if subject i links to the blue node in round t. Table A3 shows the results of a linear probability model regression of $b_{i,t}$on $b_{i,t-1}$ in the simultaneous treatment, conditional on a star network *not* being formed in either round t or $t-1$, by parameter set. The existence of early adopters is characterized by autocorrelation of $b_{i,t}$: an early adopter links to the blue node in both round t and $t-1$. The evidence is strongly in favor of the existence of early adopters: subjects in all treatments (excluding $B = 12, H = 0$) are willing to repeatedly link to the blue node in rounds where an efficient star has not formed. The alternative, that subjects independently randomize each period, is rejected by the data. The autocorrelation in $b_{i,t}$ is clearly increasing in both H and B,with the correlation being highly significant for the $B = 24$ parameterizations.

[TABLE A3 ABOUT HERE]

A natural follow-up is to determine whether subjects who persistently link to P0 are more likely to become the central player in a star network if a star network forms. That is, do early adopters facilitate the formation of star networks around themselves? The answer is yes. To establish this, we define $\bar{b_i}$ to be the proportion of rounds that player i links to the blue node in rounds prior to the first formation of a star network and c_i to be a binary variable that is equal to 1 if player i centers the first star that forms (and equal to 0 otherwise). We perform a logit regression of c_i on $\bar{b_i}$ clustering standard errors at the group level and dropping groups for which a star never forms, and report the marginal effect of $\bar{b_i}$ on c_i, calculated at the mean of $\bar{b_i}$ in table A4. Subjects who link to the blue node more often prior to the formation of a star network are more likely to be the center of a star network once it forms.

[TABLE A4 ABOUT HERE]

This analysis establishes that there exists some subjects who persistently link to P0 prior to a star being formed. Further, players who link to P0 more often are more likely to become the center of a star network in later rounds. These *early adopters* facilitate the formation of star networks. In the next section we propose a diagnostic treatment that investigates the underlying motivations of *early adopters*.

B.3 Simultaneous results in the forced linking treatment

We use the simultaneous rounds as a test that the new treatment has not substantially altered the strategic nature of our game. We expect efficient star networks to occur more often in the *forced linking* treatment because only three, rather than four, players need to coordinate their behavior to generate a star network. The proportion of star networks should increase both for rounds where there was, and was not, coordination in the previous round. However, given our sample size (32 subjects across 2 sessions), we may not be able to identify small increases in coordination and therefore hypothesize that the proportion of efficient stars should not be lower in the *forced linking* treatment than in the *baseline* treatment. Figures A2 and A3 demonstrate this to be the case. There is a decrease in the proportion of star networks, conditional on coordination in the previous round, but we cannot reject the null hypothesis that the proportion is the same across treatments (p=0.128, one-tailed test).

[FIGURE A2 ABOUT HERE]

[FIGURE A2 ABOUT HERE]

C Online appendix

Consider a set of N agents, $\{1, \ldots, N\}$. Each agent resides on exactly one node of an empty network with $N+1$ nodes, and no two agents reside on the same node. The node that player $i \in \{1, \ldots, N\}$ resides on is denoted by node-i. The empty node (source of the public good) is referred to as P0. The Network Dilemma is a network formation game played by this set of N agents. Each agent may form up to N directed edges; all edges formed by agent i must originate from node-i and terminate at a node-j where $j \in \{0, 1, \ldots, N\} \setminus i$. Given that an agent may either form or not form a link to each node, the action set for each agent has size 2^N. A network $\mathcal{N}$ is a $(N+1) \times (N+1)$ binary matrix where an entry of 1 in position (i, j) indicates a directed link from node $(i-1)$ to node $(j-1)$. The first row of any network $\mathcal{N}$ is a row of zeros, as P0 cannot form any links.

The Network Dilemma game has the following features:

- Link formation is costly: an agent who forms t links to non-P0 nodes must pay a link formation cost of $C(t)$ with $C(0) = 0$ and $C(t+1) > C(t)$ for all t.
- Linking to P0 is particularly costly: an agent who links to node-0 must pay a cost of S where $S > C(j+1) - C(j)$ for all $j \geq 0$.
- An agent benefits from being linked to P0, earning $G(p)$ when the shortest directed path from the agent to P0 is of length p with $G(p+1) \leq G(p)$ and $G(p) \geq 0$ for all $p \geq 1$. If no path exists from the agent to P0, then we write $p = \infty$ and define $G(\infty) = 0$.
- Finally, an agent benefits from being linked to by other agents: for every link that terminates at an agent's node, they earn a fixed benefit $H \geq 0$ with $H < C(j+1) - C(j)\ \forall j$.

The last two features capture the feature of localized network benefits. If the network $\mathcal{N}$ is formed, then agent i will earn payoff:

$$u_i(\mathcal{N}) = G(p_i) + m_i H - C(t_i) - s_i S$$

where p_i is the shortest (directed) path from node-i to node-0, m_i is the number of edges that terminate at node-i, t_i is the number of edges that originate at node-i and terminate at a node other than node-0 and s_i is an indicator that equals to 1 if there is an edge from node-i to node-0.

Finally, we impose two parametric assumptions that generate the dilemma structure of the Network Dilemma. First, we assume that linking toP0 is valuable enough that linking to P0 is preferred to remaining unlinked $G(1) > S$. This assumption also ensures that the empty network is never an equilibrium. Second, we assume thatlinking to P0 directly is sufficiently costly such that $G(2) - C(1) > G(1) - S$: free-riding off another player is preferred to linking to P0 directly.

We consider two timing protocols for this game. In the first, the game is played simultaneously. In the second, the game is played sequentially as a game of perfect information. The agent residing on node-1 moves first, then the agent residing on node-2, and so on. We begin by classifying the set of strict Nash equilibria of the (simultaneous) game.

Proposition 10. *In any strict Nash Equilibrium of the Network Dilemma, each agent forms exactly one edge.*

Having at least one link is always utility maximizing, given the utility flow from an optimal link to P0 is no smaller than $(G(1) - S) > 0$. No agent wishes to form more than one link because all links are costly but only the link on the shortest path to P0 is consequential for one's utility. This provides the justification for restricting subjects to forming exactly one link in our experimental implementation of the Network Dilemma. For the next proposition, we define a star network to be any network where (i) every player forms only one edge, and (ii) a single player, say i, connects to P0 and every other player connects to player i.

Proposition 11. *The aggregate utility is maximized by the star network.*

Proposition 11 characterizes payoff efficiency in the network.The payoff-optimal distance from P0 is pinned down by the condition $G(2) - C(1) > G(1) - S$ and the assumption that $C(.)$ is increasing and $G(.)$ is decreasing in path-length.[36] Then, Proposition 11 follows from the fact that a single agent connecting to the outside node is sufficient for all other agents to form a path of length 2 to P0. The network centrality benefit, $H < C(j+1) - C(j)\ \forall j$, is small enough that any additional links reduce welfare.

Proposition 12. *A network constitutes a pure strategy Nash equilibrium if and only if it is a star network.*

In a star network there are no incentives for deviation. Each player on the periphery is earning $G(2) - C(1)$, which is the maximal payoff possible for a player with no incoming links. The player in the center of the star is earning $G(1) - S + (N-1)H$, and has no profitable deviation. Adding an extra link incurs additional costs for no additional benefit, and removing their link to P0 reduces their payoff more than the cost of the link.

In the sequential formulation of the game, we restrict attention to subgame perfect Nash equilibrium (Proposition 13).

Proposition 13. *If $(N-1)H < (G(2) - C(1)) - (G(1) - S)$ then the unique subgame perfect Nash equilibrium outcome is for players 1 through $N-1$ to link to player N, and for player N to link to P0. If $(N-k)H \geq G(2) - C(1) - (G(1) - S)$ for $N > k \geq 1$ then there exists multiple subgame perfect Nash equilibrium. In each of the subgame perfect Nash equilibrium outcomes one of the first k players links to P0 and all other players link to the player who linked to P0.*

[36] Given $H < C(j+1) - C(j)\ \forall j$, links are inefficient unless they are essential for the smallest path from an agent to P0.

In the final subgame, when player N has to move, their best response is linking to P0 if no-one has previously linked to P0, and otherwise link to someone who has linked to P0. Working backwards, when H is small, the second to last player will link to the last player if no-one has previously linked to P0, and otherwise link to someone who has linked to P0. On the equilibrium path, everyone links to the last player who then links to the blue node. If the externality benefits, or N, are large enough such that the central node of a star network earns more than the periphery, then there exists multiple subgame perfect equilibria. Otherwise, there is a unique equilibrium that generates a star network centered around the last player to move.

C.1 Proofs

Proof. [Lemma 10]Suppose that there exists a pure strategy Nash equilibrium agent i has formed no edges. The payoff for agent i is therefore m_iH. If the agent formed an edge to P0 they would earn $G(1) - S + m_iH > m_iH$.

The next part of the proof is by contradiction. Suppose that there exists a pure strategy Nash equilibrium where agent j has formed multiple edges. Select an edge that is part of the shortest path to P0. By definition this path cannot include any other edges emanating from j. Deleting any other edge does not increase the least distance from P0, and hence, will increase the payoff of player j. □

Proof. [Lemma 11] A network where there are no edges terminating at P0 generates a non-positive aggregate utility. Therefore the aggregate utility maximizing network must have at least one connection to P0. Ignoring centrality benefits, the maximal individual utility available is $G(2) - C(1)$, and the links formed by such a player generate an externality H. Therefore the maximal feasible aggregate utility is $G(1) - S + (N - 1)(G(2) - C(1) + H)$. This is the payoff generated by the star network. Given that $H < C(j + 1) - C(j)$ adding additional links decreases aggregate welfare. □

Proof. [Lemma 12] Suppose that there exists a NE where no player is connected to P0. There is a profitable deviation for each player to connect to P0.

Suppose that there exists a NE where more than one player is connected to P0, each earning $G(1) - S$. There is a profitable deviation for one of those players to delete their current edge and instead connect to another player who is connected to P0, therefore earning $G(2) - C(1) > G(1) - S$.

If exactly one player, say player i, is connected to P0, then the best response for all other players is to connect to the P0-connected player. Given that no-one else is connected to P0, the best response for player i is to connect to P0. □

Proof. [Lemma 13] Let m_0^i denote the number of links made to P0 by players who moved before player i. We shall write $l_i = j$ to denote the strategy of player i linking to player j.

Case 1. *[$G(1) - S + (N-1)H < G(2) - C(1)$] We proceed by backwards induction.*
Player N: If $m_0^N = 0$, then $l_N = 0$. If $m_0^N > 0$, then $l_N = j$ where $l_j = 0$.
Player $N-1$: If $m_0^{N-1} = 0$, then $l_{N-1} = N$. If $m_0^{N-1} > 0$ then $l_{N-1} = j$ where $l_j = 0$.
Player $N-t$: If $m_0^{N-t} = 0$, then $l_{N-t} = N$. If $m_0^{N-t} > 0$ then $l_{N-t} = j$ where $l_j = 0$ for $2 \leq t \leq N-2$.
Player 1: $l_1 = N$.

Case 2. *[$G(1) - S + (N-k)H \geq G(2) - C(1)$] We construct a subgame perfect equilibrium where player $p \leq k$ forms the center of a star.*
Player N: If $m_0^N = 0$, then $l_N = 0$. If $l_p = 0$ then $l_N = p$, else $l_N = j$ where $j = \min_{q \in \mathbb{N}}\{q : l_q = 0\}$.
Player $N-1$: If $m_0^{N-1} = 0$, then $l_{N-1} = N$. If $l_p = 0$ then $l_N = p$, else $l_N = j$ where $j = \min_{q \in \mathbb{N}}\{q : l_q = 0\}$.
Player k: If $m_0^k = 0$, then $l_k = 0$. If $l_p = 0$ then $l_k = p$, else $l_k = j$ where $j = \min_{q \in \mathbb{N}}\{q : l_q = 0\}$.
Player $k-t$: If $m_0^{k-t} = 0$, then $l_{k-t} = 0$.If $l_p = 0$ then $l_{k-t} = p$, else $l_{k-t} = j$ where $j = \min_{q \in \mathbb{N}}\{q : l_q = 0\}$.
Player p: $l_{N-p} = 0$.
Player $p-t$: $l_{p-t} = p$ for $1 \leq t \leq (p-1)$.

□

References

James Andreoni. Giving with impure altruism: Applications to charity and ricardian equivalence. *Journal of Political Economy*, 97(6):1447–1458, 1989. 2.3

James Andreoni, Paul M. Brown, and Lise Vesterlund. What makes an allocation fair? some experimental evidence. *Games and Economic Behavior*, 40:1–24, 2002. 6

Pierpaolo Battigalli. On rationalizability in extensive games. *Journal of Economic Theory*, 74: 40–61, 1997. 2.1, 2.2

Ted Bergstrom, Rodney Garrat, and Greg Leo. Let me, or let george? motives of competing altruists. University of California at Santa Barbara, Economics Working Paper Series, December 2015. 18

Gary E. Bolton and Axel Ockenfels. Erc: A theory of equity, reciprocity, and competition. *American Economic Review*, 90(1):166–193, 2000. 2.1, 2.1

Bloomberg Businessweek. Forced or not? Why u.s. says china steals technology, November 2019. URL `https://www.bloomberg.com/news/articles/2019-06-15/forced-or-not-why-u-s-says-china-steals-technology-quicktake`. Accessed February 12, 2020. 4

Jane Cai and Keegan Elmer. Is the us right to cry foul about forced technology transfer to do business in china – and what is beijing's position? South China Morning Post, January 2019. URL `https://www.scmp.com/news/china/diplomacy/article/2181528/us-right-cry-foul-about-forced-technology-transfer-do-business`. Accessed February 12, 2020. 4

Jeffrey Carpenter, Shachar Kariv, and Andrew Schotter. Network architecture, cooperation and punishment in public good experiments. *Review of Economic Design*, 16(2-3):93–118, 2012. 1, 6

Gary Charness, Francesco Feri, Miguel A Meléndez-Jiménez, and Matthias Sutter. Experimental games on networks: Underpinnings of behavior and equilibrium selection. *Econometrica*, 82(5): 1615–1670, 2014. 1, 6

Gary B. Charness and Matthew Rabin. Understanding social preferences with simple tests. *Quarterly Journal of Economics*, 117(3):817–869, 2002. 2.1

Daniel L. Chen, Martin Schonger, and Chris Wickens. otree – an open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9:88–97, 2016. 19

Syngjoo Choi, Schachar Kariv, and Edoardo Gallo. Networks in the laboratory. In Yann Bramoulle, Andrea Galeotti, and Brian Rogers, editors, *The Oxford Handbook of the Economics of Networks*. Oxford, 2016. 6

Syngjoo Choi, Sanjeev Goyal, and Frédéric Moisan. Network formation in large groups. 2019. 6

Jurgen Eichberger and David Kelsey. Strategic complements, substitutes, and ambiguity: The implications for public goods. *Journal of Economic Theory*, 106(2):436–466, 2002. 1

Armin Falk and Michael Kosfeld. It's all about connections: Evidence on network formation. *Review of Network Economics*, 11(3), 2012. 1, 6

Ernst Fehr and Klaus M. Schmidt. A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics*, 114(3):817–868, 1999. 2.1, 2.1

Lawrence F. Feick and Linda L. Price. The market maven: A diffuser of marketplace innovation. *Journal of Marketing*, 51(1):83–97, January 1987. 1

Andrew D. Foster and Mark R. Rosenzweig. Learning by doing and learning from others: Human capital and technical change in agriculture. *Journal of Political Economy*, 103(6):1176–1209, December 1995. 1

Simon Gachter, Daniele Nosenzo, Elke Renner, and Martin Sefton. Sequential vs. simultaneous contributions to public goods: Experimental evidence. *Journal of Public Economics*, 94:515–522, 2010. 6

Andrea Galeotti and Sanjeev Goyal. The law of the few. *American Economic Review*, 100:1468–1492, September 2010. 9, 10, 6

Jacob K. Goeree, Arno Riedl, and Aljaz Ule. In search of stars: Network formation among heterogenous agents. *Games and Economic Behavior*, 67(2):445–466, 2009. 1, 6

Sanjeev Goyal, Stephanie Rosenkranz, Utz Weitzel, and Vincent Buskens. Information acquisition and exchange in social networks. *The Economic Journal*, 127:2302–2331, November 2017. 1, 6

Luigi Guiso and Fabiano Schivardi. Spillovers in industrial districts. *The Economic Journal*, 117 (516):68–93, 01 2007. 1

Adam B. Jaffe, Manuel Trajtenberg, and Rebecca Henderson. Geographic localization of knowledge spillovers as evidenced by patent citations. *The Quarterly Journal of Economics*, 108(3):577–598, August 1993. 1

Paul Mozur. Inside a heist of american chip designs, as china bids for tech power. New York Times, June 2018. URL `https://www.nytimes.com/2018/06/22/technology/china-micron-chips-theft.html`. Accessed February 12th 2020. 2

Prithvijit Mukherjee. Cost sharing in public goods game in networks. 2019. 1, 6, 28

Rong Rong and Daniel Houser. Growing stars: A laboratory analysis of network formation. *Journal of Economic Behavior and Organization*, 117:380–394, 2015. 1, 6

Stephanie Rosenkranz and Utz Weitzel. Network structure and strategic investments: An experimental analysis. *Games and Economic Behavior*, 75(2):898–920, 2012. 1, 6

Nidhi Subbaraman. Harvard chemistry chief's arrest over china links shocks researchers. *Nature*, February 2020. doi: 10.1038/d41586-020-00291-2. URL https://www.nature.com/articles/d41586-020-00291-2. DOI:. 1

Boris van Leeuwen, Theo Offerman, and Arthur J.H.C Schram. Competition for status creates superstars: An experiment on public good provision and network formation. *Journal of the European Economic Association*, Forthcoming. 1, 6, 29

D Tables

		Order 1	Order 2	Order 3	Order 4
Block 1	Simultaneous	$B = 12, H = 0$	$B = 12, H = 4$	$B = 24, H = 0$	$B = 24, H = 4$
Block 2		$B = 12, H = 4$	$B = 12, H = 0$	$B = 24, H = 4$	$B = 24, H = 0$
Block 3	Sequential	$B = 12, H = 4$	$B = 12, H = 0$	$B = 24, H = 4$	$B = 24, H = 0$
Block 4		$B = 12, H = 0$	$B = 12, H = 4$	$B = 24, H = 0$	$B = 24, H = 4$

Table 1: Order of blocks. Two sessions of each ordering were conducted.

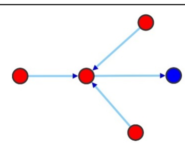
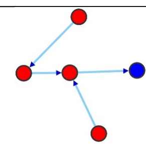
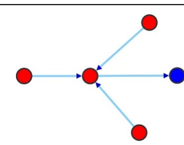
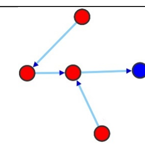
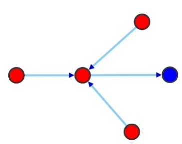
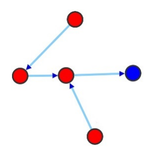
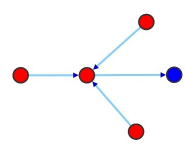
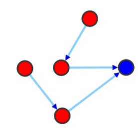
	H=0		H=4	
	Most frequent	2nd most frequent	Most frequent	2nd most frequent
$B = 12$				
Occurrences	157	76	289	48
Stability	68%	28%	76%	36%
Strict Stability	19%	0%	76%	8%
$B = 24$				
Occurrences	321	29	300	76
Stability	88%	7%	92%	60%
Strict Stability	80%	7%	92%	45%

Table 2: The two most common networks for each of the sequential treatments. Stability measures the proportion of times the same network is observed in the subsequent round. Strict stability requires, in addition, the same players to occupy the same location in the network. There are 400 total observations per treatment.

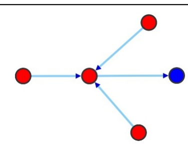
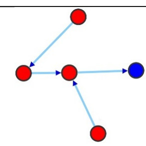
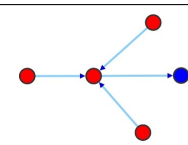
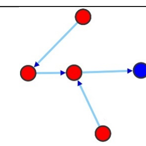
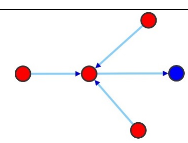
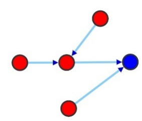
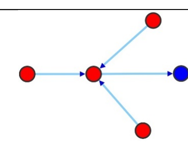
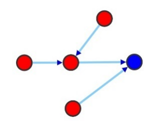
	H=0		H=4	
	Most frequent	2nd most frequent	Most frequent	2nd most frequent
B=12				
Occurrences	74	45	101	51
Stability	70%	11%	85%	24%
Strict Stability	67%	0%	85%	14%
B=24				
Occurrences	142	58	149	68
Stability	85%	25%	87%	33%
Strict Stability	85%	15%	87%	29%

Table 3: The two most common networks for each of the simultaneous treatments. Stability measures the proportion of times the same network is observed in the subsequent round. Strict stability requires, in addition, the same players to occupy the same location in the network. There are 400 total observations per treatment.

	$B = 12$	$B = 12$	$B = 24$	$B = 24$
	$H = 0$	$H = 4$	$H = 0$	$H = 4$
P1, P2 and P3 link to P4	53	46	70	0
P4 Links to P0	25	43	61	-
Proportion to P0	0.47	0.93	0.87	-
Proportion of P4 that never face dilemma	0.06	0.63	0.31	1

Table 4: Number of rounds where P1, P2 and P3 all link to P4, and number of rounds where P4 links to P0 when all other players link to P4. The proportion of P4 who never face dilemma is the proportion of groups that never, in any round, had P1, P2 and P3 all link to P4.

	SPNE-strategy payoff	Average deviation payoff	Proportion of Deviation
		$B = 12, H = 0$	
P3	19.4	11.7	0.32
P2	17.7	12.9	0.33
P1	17.4	15.7	0.71
		$B = 12, H = 4$	
P3	28.8	13.6	0.10
P2	27.7	17.1	0.33
P1	23.8	23.7	0.81
		$B = 24, H = 0$	
P3	27.4	21.2	0.13
P2	26	22.2	0.13
P1	24.1	23.7	0.77

Table 5: Average empirical payoffs and the proportion of deviations from the SPNE by player in the sequential treatment, conditional on all previous movers playing the SPNE (i.e. P2 payoffs and behavior are conditional on P1 linking to P4, and P3 payoffs and behavior are conditional on P1 and P2 linking to P4).

	B=12		B=24	
	H=0	H=4	H=0	H=4
P1	$16.2^{2,(3),(4)}$	$23.7^{(3),4}$	$23.8^{(2),(3),(4)}$	$33.5^{(2),(3),(4)}$
P2	$18.9^{1,(3),(4)}$	25.7^{3}	$27.1^{(1)}$	$29.8^{(1)}$
P3	$21.1^{(1),(2)}$	$27.7^{(1),2}$	$27.9^{(1)}$	$29.8^{(1)}$
P4	$20.8^{(1),(2)}$	27.2^{1}	$27.8^{(1)}$	$29.9^{(1)}$

Table 6: Average payoffs by subject role in the sequential treatment. A superscript j on an entry in row i indicates that the payoff difference between Pj and Pi is statistically significant at the 5% level; a superscript (j) indicates significance at the 1% level. Standard errors are clustered at the group level.

P1 links to	Possible outcomes	Payoff
Pi $i \geq 2$	Players $j \in \{1, i\}$ link to Pi, Pi links to any node	$\pi_i = B + 3H$ $\pi_{-i} = 30$
	All other players link to P1	$\pi_1 = B + 3H$ $\pi_{-1} = 30$
P0	All other players link to P1	$\pi_1 = B + 3H$ $\pi_{-1} = 30$

Table 7: Rationalizable outcomes and payments in the forced linking treatment.

	No star	P1 star	P2 star	P3 star	P4 star
$\mathbb{1}(H = 4, B = 12)$	0.098	-0.100	0.100	-0.100	-0.002
	(0.056)	(0.095)	(0.060)	(0.061)	(0.091)
$\mathbb{1}(forcedlinking)$	-0.125^{**}	-0.244^{**}	-0.054	-0.015	0.438^{***}
	(0.050)	(0.092)	(0.056)	(0.066)	(0.109)
Constant	0.188^{***}	0.481^{***}	0.076^{**}	0.123	0.131
	(0.027)	(0.072)	(0.029)	(0.063)	(0.071)
N	1200	1200	1200	1200	1200
No. of clusters	48	48	48	48	48

Table 8: Linear regression of the probability of network outcomes on treatment and payoff parameters. The ommitted category is the *baseline* treatment with $H = 0, B = 24$. Standard errors, clustered at the group-block level, are in parenthesis. $^{**}: p < 0.05$, $^{***}: p < 0.01$

	Connect to P0	Connect to non-P0
$B = 12, H = 0$	$\frac{12}{102} = .118$	$\frac{10}{40} = .250$
$B = 12, H = 4$	$\frac{24}{114} = .210$	$\frac{22}{56} = .392$
$B = 24, H = 0$	$\frac{24}{114} = .210$	$\frac{10}{40} = .250$
$B = 24, H = 4$	$\frac{36}{126} = .29$	$\frac{22}{56} = .392$

Table A1: P4 payment as a proportion of total payments

	Coord_Idx	Pr(star|star)	Pr(star|not star)
$\mathbb{1}(H = 4)$	0.887***	0.096***	0.009
	(0.204)	(0.034)	(0.028)
$\mathbb{1}(B = 24)$	1.647***	0.093**	0.073**
	(0.204)	(0.037)	(0.028)
Sequential	1.349***	0.037	0.217***
	(0.204)	(0.05)	(0.04)
Const.	−0.781***	0.722***	0.054**
	(0.241)	(0.051)	(0.022)
N	128	1148	1614
No. of clusters	128	110	123

Table A2: OLS regression of the coordination index, and the probability of star formation conditional on whether a star was observed in the previous period, on treatment parameters. The omitted group is the simultaneous treatment with $B = 12, H = 0$. Standard errors are in brackets and clustered at the group level. All independent variables treated as categorical. $^{**}: p < 0.05$, $^{***}: p < 0.01$

	$b_{i,t}$ ($B=12, H=0$)	$b_{i,t}$ ($B=12, H=4$)	$b_{i,t}$ ($B=24, H=0$)	$b_{i,t}$ ($B=24, H=4$)
$b_{i,t-1}$	-0.064	0.123**	0.222***	0.490***
	(0.036)	(0.058)	(0.052)	(0.064)
Const.	0.212***	0.244***	0.360***	0.264***
	(0.020)	(0.023)	(0.035)	(0.044)
N	1156	1076	884	868
No. of clusters	64	64	64	64

Table A3: Regression of the probability of linking to the blue node on the probability of linking to the blue node in the previous round, conditional on a star network not being formed in either round. Simultaneous treatment. Standard errors, clustered at the individual level, are in brackets. $^{**}: p < 0.05$, $^{***}: p < 0.01$

	$B=12$		$B=24$	
	$H=0$	,$H=4$	$H=0$	$H=4$
Mean of $\bar{b_i}$	0.26	0.35	0.45	0.57
Marginal effect at mean	1.12***	0.95***	0.63**	0.56***
	(0.29)	(0.36)	(0.29)	(0.19)
N	48	48	52	52
No. of clusters	12	12	13	13

Table A4: Marginal effect of the proportion of rounds a subject links to the blue node prior to a star network forming on the probability that the subject becomes the center of a star network. Marginal effects calculated using logit regression with standard errors clustered at the group level. $^{**}: p < 0.05$, $^{***}: p < 0.01$

B Figures

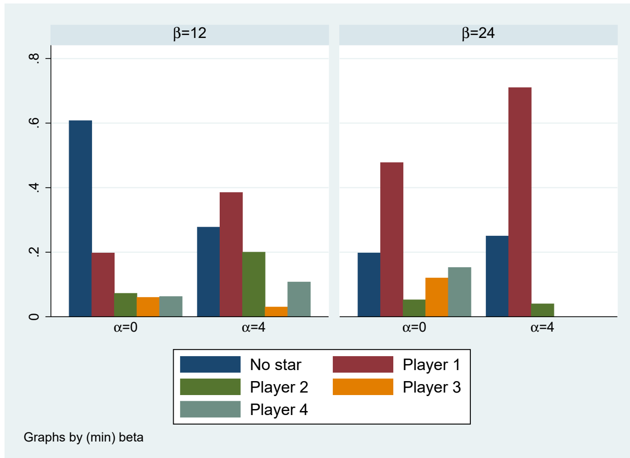


Figure 1: Probability of forming an efficient star network, by central player, in the sequential treatment.

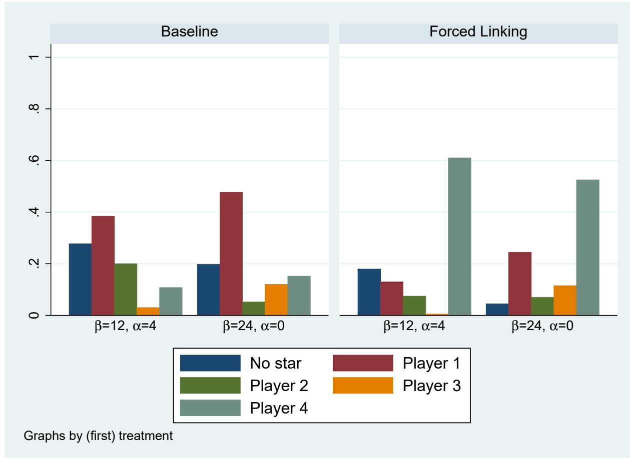


Figure 2: Probability of forming an efficient star network, by central player, in the sequential treatment, split by *baseline* and *forced linking*.

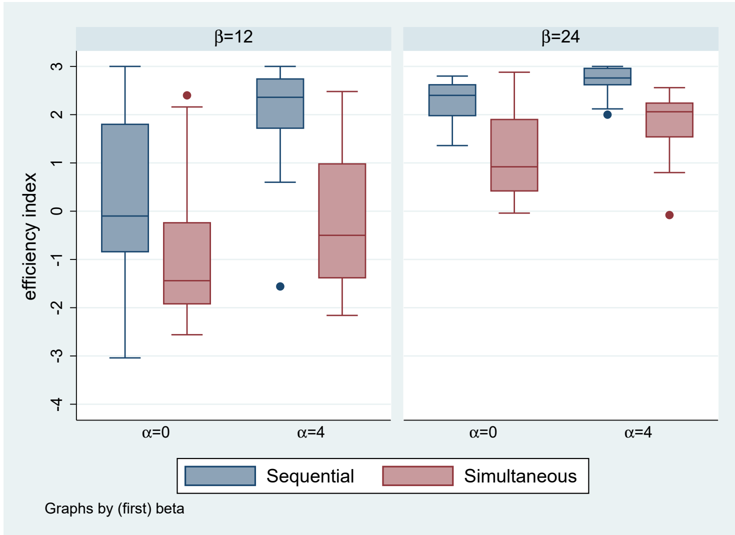


Figure A1: Coordination index by treatment.

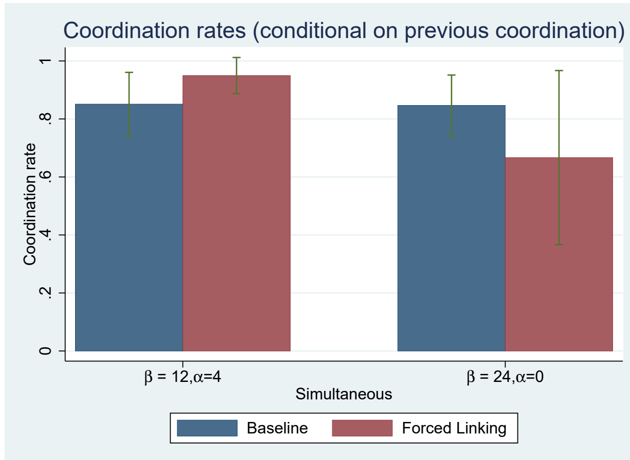


Figure A2: Probability of forming an efficient star network, conditional on having formed an efficient star in the previous round, in the simultaneous rounds with 95% confidence intervals.

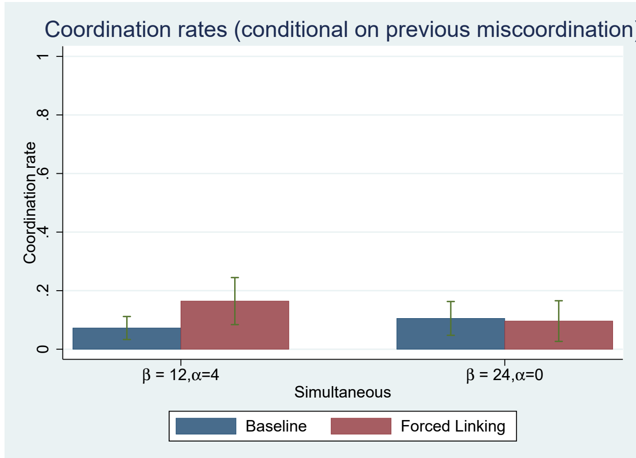


Figure A3: Probability of forming an efficient star network, conditional on not having formed an efficient star in the previous round, in the simultaneous rounds with 95% confidence intervals.