UNIVERSITY OF CAMBRIDGE
Faculty of Economics

Institute for New Economic Thinking

Cambridge-INET Institute

Cambridge-INET Working Paper Series No: 2017/08

Cambridge Working Papers in Economics: 1721

INTEGRATION AND DIVERSITY

Sanjeev Goyal
(University of Cambridge)

Penélope Hernández
(Universitat de Valéncia)

Guillem Martínez-Cánovasz
(Universitat de Valéncia)

Frédéric Moisan
(University of Cambridge)

Manuel Muñoz-Herrera
(Universitat de Valéncia)

Angel Sánchez
(Universidad Carlos III de Madrid
Universidad de Zaragoza)

(revised April 2020)

We study a setting where individuals prefer to coordinate with others but they differ on their preferred action. Our interest is understanding the role of linking in shaping behavior. So we consider the situation in which interactions are exogenous and a situation where individuals choose links that determine the interactions. Theory is permissive in both settings: conformism (on either of the actions) and diversity (with different groups choosing their preferred actions) are both sustainable in equilibrium. Our experiments reveal that, in an exogenous complete network, subjects choose to conform to the majority's preferred action. By contrast, when linking is free and endogenous, subjects form dense networks (biased in favour of linking within same preferences type) but choose diverse actions. The convergence to diverse actions is faster under endogenous linking as compared to the convergence to conformity on the majority's preferred action under the exogenous complete network. Thus our experiments suggest that individuals use links to resolve the coordination problem.

Integration and Diversity

Sanjeev Goyal[*] Penélope Hernández[†] Guillem Martínez-Cánovas[‡]
Frédéric Moisan[§] Manuel Muñoz-Herrera[¶] Angel Sánchez[‖]

April 13, 2020

Abstract

We study a setting where individuals prefer to coordinate with others but they differ on their preferred action. Our interest is in understanding the role of link formation with others in shaping behavior. So we consider the situation in which interactions are exogenous and a situation where individuals choose links that determine the interactions. Theory is permissive in both settings: conformity (on either of the actions) and diversity (with different groups choosing their preferred actions) are both sustainable in equilibrium.

Our experiments reveal that, in an exogenous complete network, subjects choose to conform to the majority's preferred action. By contrast, when linking is free and endogenous, subjects form dense networks (biased in favour of linking within same preferences type) but choose diverse actions. The convergence to diverse actions is faster as compared to the convergence to conformity. Thus, our experiment suggests that endogenous links facilitate quick resolution of the coordination problem.

JEL Codes: D85, D03, C72, C92

Keywords: Networks, equilibrium selection, social coordination, experiments

[*]Faculty of Economics and Christ's College, University of Cambridge. Email: sg472@cam.ac.uk

[†]Departamento de Análisis Económico, Universitat de València, Email: penelope.hernandez@uv.es.

[‡]Departamento de Análisis Económico, Universitat de València, Email: guillem.martinez@uv.es.

[§]Faculty of Economics, University of Cambridge. Email: fm442@cam.ac.uk

[¶]Social Science Division, New York University Abu Dhabi, Email: manumunoz@nyu.edu

[‖]Departamento de Matemáticas, Universidad Carlos III de Madrid. Email: anxo@math.uc3m.es.

Acknowledgements:

We are grateful to the editor and two anonymous referees for comments that have improved the paper. This paper has been supported by the EU through FET-Proactive Project DOLFINS (contract no. 640772) and FET-Open Project IBSEN (contract no. 662725), grant FIS2015-64349-P (MINECO/FEDER, UE), and grant ECO2017-87245-R (MINECO/FEDER, UE). We are grateful to Marina Agranov, Gary Charness, Vince Crawford, Sihua Ding, Matt Elliott, Edoardo Gallo, Joerg Kalbfuss, Jonathan Newton, Theo Offerman, Gustavo Paez, Romans Pancs, Debraj Ray, Arno Reidl, Marzena Rostek, Robert Sugden, Alan Walsh, Sevgi Yuksel, and participants at a number of seminars for helpful comments.

1 Introduction

Predicting which of the many equilibria will be selected is perhaps the most difficult problem in game theory (Camerer 2003)

Consider a community with two cultural groups, a native population with its language and an immigrant group, with its own language and distinct cultural background. Individuals prefer to use their own language as they are emotionally attached to it (and also because they are more fluent in it). At the moment, families can decide where they wish to live. In recent years, however, the immigrant group has grown and there is a fear that this could lead, over time, to fragmentation of the community into distinct neighbourhoods (corresponding to the two cultural groups). The community is considering a proposal of quotas in large residential complexes. Should the community fix such quotas? The goal of the present paper is to shed light on questions of this sort.

To clarify the key considerations, we start by setting out a theoretical model. There is a group of individuals who each choose between two actions: *up* or *down*. Everyone prefers to coordinate on one action but individuals differ in the action they prefer: group U prefers action *up*, group D prefers action *down*. We consider a baseline setting in which everyone is obliged to interact with everyone else, and a setting in which individuals choose with whom to interact. In the latter setting, everyone observes the network that is created and then chooses between action *up* and *down*. The theoretical analysis reveals a rich set of possibilities.

Consider the case where everyone interacts with everyone else.[1] There exist three equilibria: everyone conforming to a single action, *up* or *down*, and diversity with everyone choosing their preferred action (i.e., group U members choosing *up*, and group D choosing *down*). Next, consider the setting with endogenous linking, and suppose that the costs of linking are zero. Now the outcomes take two forms: one, every individual connects to everyone else and the action profile corresponds to the three equilibria described above. The other situation exhibits partial connectivity: an interesting special case arises when

[1] Formally, we refer to this as the exogenous complete network.

the network fragments into two distinct components and individuals in each component choose a different action. Moreover, we show that in both the exogenous and endogenous interaction setting, conforming to the majority's preferred action maximizes aggregate welfare.[2] Thus, there are multiple equilibrium outcomes, in both the exogenous and the endogenous linking case, and there is a tension between diversity and aggregate welfare.[3] We conduct laboratory experiments to better understand how players choose actions and how these choices are affected by whether the network is exogenous or endogenous.

The experiments involve groups of 15 subjects who play the game repeatedly, over 20 rounds. In each group, there is a majority sub-group with 8 subjects (who prefer action *up*) and a minority sub-group with 7 subjects (who prefer action *down*). We find that, with exogenous interaction, conformity on the majority's preferred action obtains in 5 out of 6 groups. By contrast, with endogenous linking, individuals form most of the possible links (roughly 95 out of a possible 105), and yet in all groups they rapidly converge on diversity. Thus, the freedom to create links has a powerful effect on behavior and on aggregate welfare.[4]

We briefly relate these findings with our motivating example. Our experiments suggest that if individual families are free to choose where to locate then it is possible that segregation will emerge in the community, whereas if members of different communities are obliged to live together – as a result of quotas – then members of the community are more likely to conform to the majority language.

To test the robustness of this role for endogenous linking, we vary the costs of linking. Different costs of linking lead to different networks: we study if the effects of endogenous linking seen with zero cost are robust to this change. We first turn to negative linking costs (or link subsidy): our next finding is that, in all the 6 groups we studied, subjects

[2]Indeed, in the experimental setting, the outcome with conformity on the majority's action Pareto dominates the outcome with diversity.

[3]In Section 6, we discuss a number of alternative equilibrium selection models.

[4]We also considered an experimental treatment with a minority of 3 members, and a majority of 12: when the minority is so small we find that the freedom to form links makes no difference. Subjects choose to conform with the majority's preferred action both in the exogenous complete network as well as when links are endogenous. This treatment is presented in Appendix C.2.

form dense (and almost complete) networks but that they choose diverse actions. Finally, we consider the case with positive costs. Our final finding is that, in all the 6 groups we studied, subjects select the outcome with segregation and diversity.

To summarize, with the complete exogenous network, subjects choose to conform on the majority's preferred action. By contrast, in all the treatments with endogenous linking – with zero, negative and positive costs of links – subjects always opt for diversity of actions. Thus diversity is a robust outcome under endogenous linking.

One reason that the diversity outcome is surprising is that the payoffs in this equilibrium are Pareto dominated by the conformism outcome. So we examine the experimental payoffs more closely. We find, somewhat surprisingly, that average minority payoffs under the exogenous complete network are *not* significantly different from the average payoffs obtained with the diversity outcome under the endogenous treatment. The main reason for this is the differential rate of convergence: minority subjects converge significantly more quickly to the steady state action profile in the endogenous linking treatment (as compared to the exogenous treatment).[5] Taking these observations together leads us to the view that endogenous linking by creating a distinct set of groups facilitates a quicker resolution of the coordination problem.

Our paper is a contribution to the study of social coordination. Following the early contributions of Schelling (1960) and Lewis (1969), there is now a large body of research on coordination problems. Blume (1993) and Ellison (1993) drew attention to the role of interaction structures in shaping coordination, while Goyal and Vega-Redondo (2005) and Jackson and Watts (2002) developed models in which players choose partners and also actions in a coordination game. In more recent years, a number of researchers have introduced heterogeneity of preferences in these models as a way to think about culture and identity, see e.g., Advani and Reich (2015), Bojanowski and Buskens (2011), and Ellwardt et al. (2016) and Neary (2012). Our paper conducts an experimental investigation on the role of endogenous linking in such a setting.

[5] Majority group subjects choose their preferred action and persist with that action from early on, in both treatments.

There is a large experimental literature on social coordination, see e.g. Charness et al. (2014), Crawford (1995), Isoni et al. (2014). Specifically, there is a strand of work on coordination games on networks (Choi and Lee 2014; Antonioni et al. 2013; Kearns et al. 2012) and a strand of work on network formation (Bernasconi and Galizzi 2005; Goeree et al. 2009).[6] We combine network formation with coordination in the current paper. An early paper by Corbae and Duffy (2008) presents an experiment where subjects form links in the first stage and play the stag hunt game in the second stage. The players in their experiment are ex-ante identical and the authors abstract from size effects, by considering average payoffs (across interactions). In our case, players belong to different preference groups (these preferences types are common knowledge) and the key tension turns to the size of interaction group. In a recent paper, Riedl et al. (2016) bring out the positive role of endogenous linking in facilitating social welfare. That paper studies linking in a minimum effort game and it finds that endogenizing the choice of partners has a dramatic effect on behavior: players converge to the most efficient Nash equilibrium. By contrast, in our paper, introducing endogenous links leads to play converging to a Pareto-dominated outcome. Thus, our work shows that endogenizing linking can have very different consequences for social welfare, depending on whether individuals have heterogeneous or similar preferences.[7]

The paper is organized as follows. Section 2 presents the model and the theoretical analysis. Section 3 presents our experimental design and Section 4 the experimental findings on endogenous versus exogenous networks. Section 5 provides alternative treatments that test the robustness of such findings. Section 6 discusses four alternative theoretical approaches — stochastic stability, team reasoning, social preferences, and k-level reason-

[6] The present paper reports an experiment with human subjects; there is also a literature that studies simulations of complex network dynamics. For a recent paper in this line of work, that studies network linking and segregation, see Lipari et al. (2019).

[7] Kearns et al. (2012) and Kearns et al. (2009) study voting behaviour by biased voters. In this game, players must coordinate on the same vote to earn a payoff. Individuals differ on their preferred outcome. Kearns et al. (2009) show that with exogenous networks subjects are quite successful in achieving coordination. By contrast, Kearns et al. (2012) show that with endogenous linking, subjects form rich networks but fail to reach coordination. This finding is in the same spirit as our work: with conflicting preferences, endogenous linking can lead to a decrease in welfare.

ing – to explain our findings. Section 7 concludes. Appendix A contains some of the proofs, Appendix B provides additional analyses of the experimental data, Appendix C contains some additional experiments while Appendix E contains the instructions for the experiments.

2 Theory

We study a game of network formation and action choice in which individuals benefit from selecting the same action as their neighbours. However, individuals differ on their preferred action. There are thus two types of individuals. We study networks that are stable and describe the corresponding equilibrium actions.

2.1 The model

Let $N = \{1, 2, \ldots, n\}$ with $n \geq 3$. The game has two stages. In the first stage, every player $i \in N$ chooses a set of link proposals g_i with others, $g_i = (g_{i1}, \ldots, g_{ii-1}, g_{ii+1}, \ldots, g_{in})$, where $g_{ij} \in \{0, 1\}$ for any $j \in N \backslash \{i\}$. Let $G_i = \{0, 1\}^{n-1}$ define i's set of link proposals. The induced network $g = (g_1, g_2, \ldots, g_n)$ is a directed graph. The closure of g is an undirected network denoted by $\overline{g}$ where $\overline{g}_{ij} = g_{ij}g_{ji}$ for every $i, j \in N$. We define the finite set of all undirected networks $\overline{g}$ as $\overline{G}$. Player i's strategy in the second stage is defined through a function x_i mapping every undirected network $\overline{g}$ that can result from the first stage to an action in $A = \{up, down\}$. Formally, $x_i : \overline{G} \to A$, and we define X_i as the set of all such strategies for player i. We denote the set of overall strategies of player i in the full game as $S_i = G_i \times X_i$, and the set of overall strategies for all players as $S = S_1 \times \ldots \times S_n$. A strategy profile $s = (x, g)$ specifies the link proposals made by every player in the first stage through $g = (g_1, g_2, \ldots, g_n)$, and the choice functions made by each player in the second stage through $x = (x_1, x_2, \ldots, x_n)$. We define $N_i(\overline{g}) = \{j \in N : \overline{g}_{ij} = 1\}$ as the set of i's neighbours in the network $\overline{g}$.

Moreover, for every player i, let $\theta_i \in \{up, down\}$ define i's type. This leads us to define $N_u = \{i \in N : \theta_i = up\}$ and $N_d = \{i \in N : \theta_i = down\}$ as the groups of players preferring action *up* and *down*, respectively ($N_u \cup N_d = N$). If $|N_u| \neq |N_d|$, we refer to the largest group of players sharing the same type/preferences as the *majority* and the other group as the *minority*. Furthermore, we define

$$\chi_i(\overline{g}, x) = \{j \in N_i(\overline{g}) : x_j = \theta_i\} \tag{1}$$

as the set of i's neighbours who play i's preferred action ($\chi_i(\overline{g}) \subseteq N_i(\overline{g})$). In what follows, we shall write $\overline{g} - \overline{g}_{ij}$ (resp. $\overline{g} + \overline{g}_{ij}$) to refer to an undirected network $\overline{g}'$ such that $\overline{g}'_{ij} = 0$ (resp. $\overline{g}'_{ij} = 1$) and $\overline{g}'_{kl} = \overline{g}_{kl}$ if $k \notin \{i, j\}$ or $l \notin \{i, j\}$.

Given strategy profile s, the utility for player i is defined as:

$$u_i(x, \overline{g}) = \lambda_{x_i}^{\theta_i}(1 + \sum_{j \in N_i(\overline{g})} I_{\{x_i = x_j\}}) - |N_i(\overline{g})|k \tag{2}$$

where $I_{x_j = x_i}$ is the indicator function of i's neighbour j choosing the same action as player i. The parameter λ is defined as follows: $\lambda_{x_i}^{\theta_i} = \alpha$ if $x_i(\overline{g}) = \theta_i$ (i chooses his preferred action), and $\lambda_{x_i(\overline{g})}^{\theta_i} = \beta$ if $x_i(\overline{g}) \neq \theta_i$ (i chooses his least preferred action) with $\beta < \alpha$. This payoff function is taken from Ellwardt et al. (2016). We note that the utility is additive across interactions with neighbours. Thus the size of the neighbourhood is an important factor in our setting.

To focus on the interesting cases, we will assume a cost of forming a link $k < \beta$. Observe that if $\beta < k$, then no player will benefit from playing their less preferred action. Moreover, if $\alpha < k$, then no player benefits from forming any link.

2.2 Equilibrium analysis

This section studies equilibrium networks and behavior. We solve backwards, starting with behavior in a given network. We then move to stage 1 and solve for stable networks.

For ease of exposition, we will drop the argument $\overline{g}$ and simply refer to strategies by x_i (instead of $x_i(\overline{g})$) whenever possible. Player i's payoff from choosing θ_i is $\alpha(|\chi_i(\overline{g})| + 1)$ and from choosing the other action is $\beta(N_i(\overline{g}) - |\chi_i(\overline{g})| + 1)$. So he is strictly better off choosing θ_i if and only if

$$\alpha(|\chi_i(\overline{g})| + 1) > \beta(|N_i(\overline{g})| - |\chi_i(\overline{g})| + 1). \tag{3}$$

This inequality can be rewritten as

$$|\chi_i(\overline{g})| > \frac{\beta}{\alpha + \beta}|N_i(\overline{g})| - \frac{\alpha - \beta}{\alpha + \beta} \tag{4}$$

Intuitively, a player is better off selecting his preferred action if and only if the proportion of his neighbours in $\overline{g}$ selecting the same action is sufficiently large. To illustrate the implications of this inequality, we consider a complete network. This network is interesting as it captures a situation of full integration where every player interacts with every other player.

Proposition 1. *Fix a complete network g. Everyone choosing the same action is an equilibrium if and only if $n \geq \alpha/\beta$. Every player choosing their preferred action is an equilibrium if and only if $|N_u|, |N_d| \geq \frac{\beta(n+1)}{\alpha+\beta}$.*

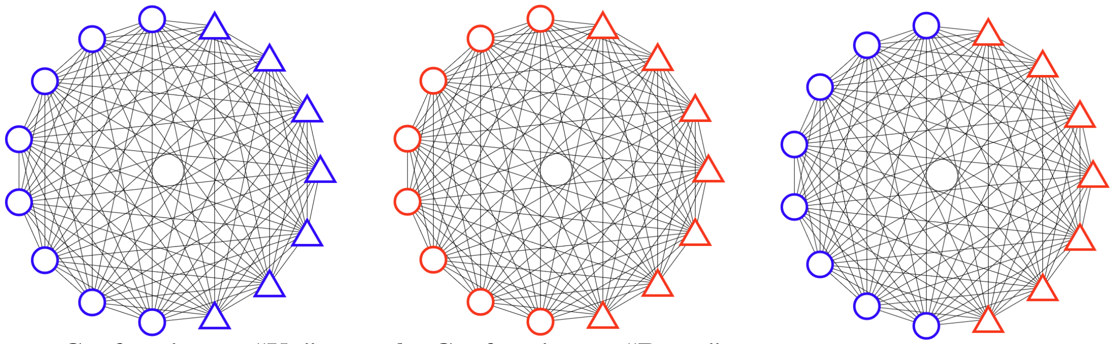
We sketch the proof here. To fix ideas, consider conformity on the majority's preferred action *up*. The payoff to a majority individual is $n\alpha$ and the payoff to a minority individual is $n\beta$. Since a deviating minority individual would obtain a payoff of α, it then follows that conformity is an equilibrium if $n \geq \alpha/\beta$. Similar computations also hold for the conformity on the minority preferred equilibrium (on action *down*).

Turning to the diversity outcome, note that if some player i benefits by playing $x_i \neq \theta_i$, then so would every player j of the same type. It then follows that diversity is an equilibrium if:

$$|N_y| - 1 \geq \frac{\beta}{\alpha + \beta}(n - 1) - \frac{\alpha - \beta}{\alpha + \beta}. \tag{5}$$

for $y \in \{u, d\}$. This inequality can be rewritten as

$$|N_y| \geq \frac{\beta(n + 1)}{\alpha + \beta} \tag{6}$$



a. Conformity on "Up" b. Conformity on "Down" c. Diversity

Figure 1. Nash equilibria in the complete network.

Note: A *circle* node represents a player in the majority and a *triangle* a player in the minority. Majority players prefer action *up* represented by color **blue**, while minority players prefer action *down* represented by color **red**. The border color of a node displays its chosen action.

for any $y \in \{u, d\}$. This completes the proof.

In a complete network there are three equilibrium outcomes: *conformity* where every player coordinates on the same action, *up* or *down*, and *diversity* where every player chooses their preferred action. Observe that conformity outcomes are always equilibria, regardless of the fraction of different types. On the other hand, the existence of the diversity outcome is contingent on a sufficiently large minority. Figure 1 illustrates these equilibrium outcomes in a society with 15 individuals divided into two types: 8 players are represented by *"circles"* and the remaining 7 individuals are represented by *"triangles"*. The circles prefer action *up*, while the triangles prefer action *down*. In all the figures throughout the article, action *up* is represented by color *"blue"* while action *down* is represented by color *"red"*.

We now solve the two stage game with link formation and action choices. We adapt the pairwise stability notion from Jackson and Wolinsky (1996) to our setting. In the spirit of their definition, we say that a network and corresponding equilibrium action profile is stable if no individual can profitably deviate either unilaterally or with one other individual. Given a network action pair $(\overline{g}, x(\overline{g}))$, $x_{-ij}(\overline{g})$ refers to the choices of all players, other than players i and j.

Definition 1. *A network-action pair* $(\overline{g}, x(\overline{g}))$ *is pairwise stable if:*

- *$x(\overline{g})$ is an equilibrium action profile given network $\overline{g}$.*

- *for every $\overline{g}_{ij} = 1$, $u_i(x, \overline{g}) \geq u_i(x', \overline{g} - \overline{g}_{ij})$ and $u_j(x, \overline{g}) \geq u_j(x'', \overline{g} - \overline{g}_{ij})$, where $x'(\overline{g} - \overline{g}_{ij})$ and $x''(\overline{g} - \overline{g}_{ij})$ are some equilibrium action profiles given network $\overline{g} - \overline{g}_{ij}$.*

- *for every $\overline{g}_{ij} = 0$, $u_i(x, \overline{g}) \geq u_i(x', \overline{g} + \overline{g}_{ij})$ or $u_j(x, \overline{g}) \geq u_j(x', \overline{g} + \overline{g}_{ij})$ where $x'(\overline{g} + \overline{g}_{ij})$ is some equilibrium action profile given network $\overline{g} + \overline{g}_{ij}$.*

In this definition, part (2) says that no player can delete an existing link and profit, while part (3) says that no pair of players can form an additional link and increase their payoffs. In both cases, note that we allow for the action profiles that would result from different networks to be independent (and therefore possibly distinct) of each other. In that sense, a single linking change can lead all players (not only those affecting the linking change) to re-optimize their actions. Our aim here is to show that conformity and diversity can both be supported in a pairwise stable outcome; moreover, these outcomes can be supported by fairly different network structures. We believe that this general observation is robust in the sense that it does not depend on specific details of the definition above.

Proposition 2. *Suppose $k = 0$. Then $(\overline{g}^*, x^*(\overline{g}^*))$ is pairwise stable if one of the following obtains:*

(i) $\overline{g}^*$ is a complete network and $x_i^*(\overline{g}^*) = m$ for all $i \in N$, where $m \in \{up, down\}$.

(ii) $\overline{g}^*$ is a complete network and $x_i^*(\overline{g}^*) = \theta_i$ for all $i \in N$, and $|N_u|, |N_d| \geq \frac{\beta n}{\alpha+\beta} + 1$.

(iii) $\overline{g}^*$ contains two complete components, $C_u = N_u$ and $C_d = N_d$ where every player in C_u chooses up, while every player in C_d chooses down.

This result highlights three types of equilibrium outcomes. Proposition 2(i) describes *Integration with conformity*, which arises when the network is complete and everyone chooses the same action. The corresponding proof follows from the fact that conformity on any action is an equilibrium for the complete network and any network with only one missing link. Since the pair of players deleting a link would earn strictly less in the subgame where they conform on the same action, the complete network is pairwise stable.

Proposition 2(ii) describes *Integration with diversity*, which arises when the network is complete and everyone chooses their preferred action. It is easy to see that such a diversity outcome is an equilibrium in any network with only one missing link between two players of different types. Moreover, such an outcome is an equilibrium in any network with only one missing link between two players of the same type if the total number of such players is sufficiently large, i.e., $|N_u|, |N_d| \geq \frac{\beta n}{\alpha+\beta} + 1$. Given every player selects their preferred action, disconnecting any pair of same type players can only decrease their payoff.

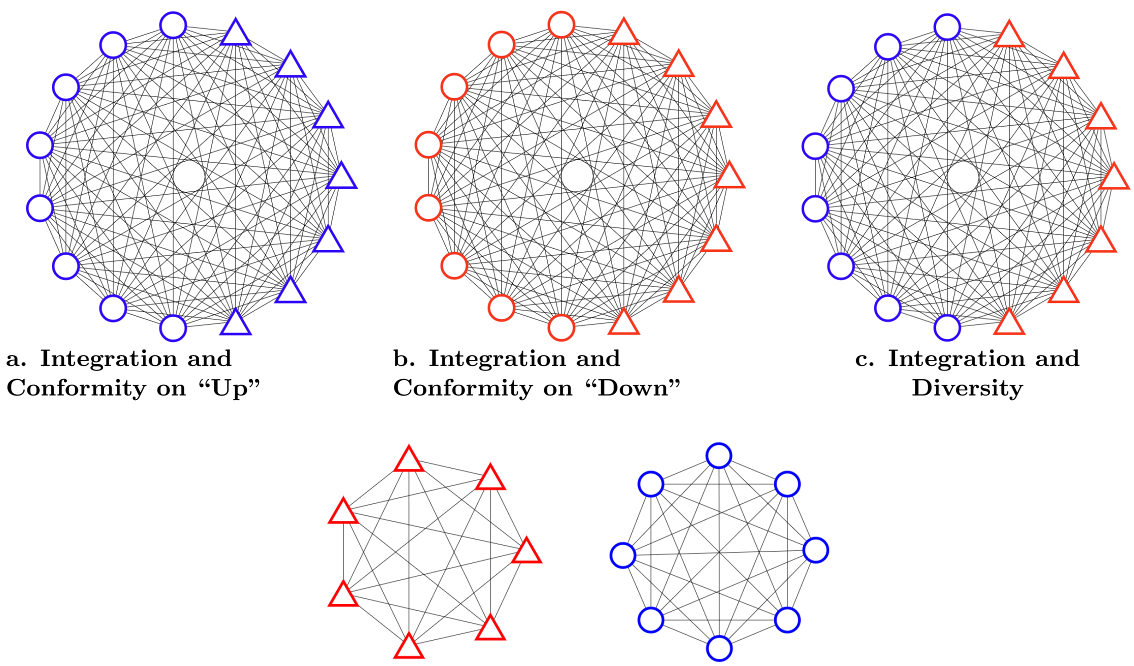
Proposition 2(iii) describes *Segregation with diversity*, which arises when the network contains two components where all individuals choose their preferred action, and members of the same component share the same type. In this case, choosing the same actions would still clearly be optimal for all if only two players of the same component were disconnected, or if two players of different types became connected with each other. Fixing the same action profile, deleting a link is clearly decreasing the corresponding players' payoffs, and adding a link has no consequences on any player's payoffs.

We illustrate these outcomes with our example ($n = 15$, $|N_{circle}| = 8$, and $|N_{triangle}| = 7$). The conformity and diversity outcomes with integration are illustrated in the top half of Figure 2, while the segregation is illustrated in the bottom half of Figure 2.

We now turn to social welfare which we define as the sum of payoffs of all players. An outcome is said to be socially efficient if it maximizes aggregate welfare. We show that both with the complete network and with endogenous networks, conformity on the majority's preferred action maximizes social surplus.

Proposition 3. *In a complete network, conformity on the majority's preferred action is socially efficient. In the game with linking and action choice, the socially efficient outcome entails a complete network and conformity on the majority's preferred action.*

The proof is presented in Appendix A. The result says that in our setting diversity is never socially desirable. To develop some intuition for the result, consider the complete network. Fixing the behavior of one group, the total aggregate payoffs can only decrease when the other group mixes actions. This follows from the coordination externalities inherent to our model. We therefore only need to compare the two outcomes: one, where



a. Integration and Conformity on "Up"

b. Integration and Conformity on "Down"

c. Integration and Diversity

d. Segregation and Diversity

Figure 2. Pairwise stable outcomes for $k = 0$.

Note: A *circle* node represents a player in the majority and a *triangle* a player in the minority. Majority players prefer action *up* represented by color **blue**, while minority players prefer action *down* represented by color **red**. The border color of a node displays its chosen action.

everyone conforms to action *up*, and two, where everyone conforms to action *down*. The concluding step then shows that conformity on *up* is better if and only if the group that prefers *up* constitutes a majority.[8] Thus with exogenous complete network, the socially efficient outcome corresponds to Figure 1(a). Similarly, in the endogenous linking treatment, the unique socially efficient outcome is presented in Figure 2(a). Note that this socially efficient outcome is invariant with respect to value of the linking cost k (so long as it is below β).

In some circumstances, we may wish to consider Pareto-domination. It is easy to see that the majority group is always better off when everyone conforms to the majority's preferred action, but the minority may or may not be better off. Assuming that the network is complete, it is easy to verify that conformity on the majority's preferred action Pareto-dominates diversity in actions if $n/min\{N_u, N_d\} > \alpha/\beta$.

We summarize the theoretical analysis as follows: in the exogenous complete network there exist multiple equilibria exhibiting conformity and diversity. The conformity equilibria exist regardless of the group size, while the diversity equilibria can only arise if the minority group is not too small. With endogenous linking, there exist multiple equilibria exhibiting full integration with conformity, full integration with diversity, and segregation with diversity. Those equilibria hold regardless of the group size. In both the exogenous complete network and the endogenous network setting, conformity on the majority's preferred action maximizes aggregate social welfare.

We now conduct laboratory experiments to examine how allowing for network formation shapes the patterns of social coordination.

[8] It is worth noting that this argument holds for arbitrary values of α and β. Thus conformity is preferred even if α is much larger than β: this is because the minority collectively gains less than what the majority losses when the minority switches away from conformism to diversity.

3 Experiment design

3.1 Basic setup

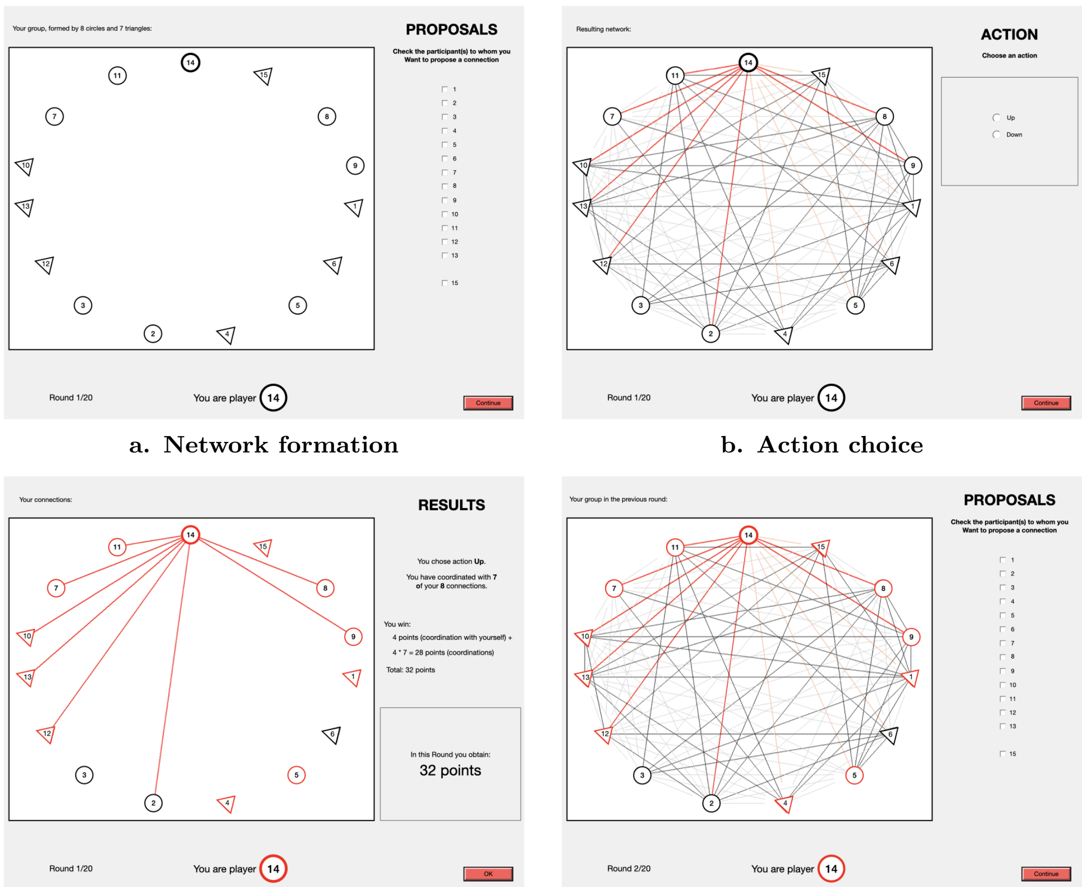
To evaluate the effects of linking on coordination and on welfare, we study two main types of treatments where we vary the way networks are formed. Treatments with *endogenous* networks start with an empty network and refer to the two stage model of linking and action choice. Treatments with *exogenous* networks specify that players are located in an exogenously given network, from which they simply choose between two coordination actions.

We consider groups of 15 participants, who interact repeatedly within the same group for 20 rounds (plus 5 unpaid trial rounds). Prior to the start of play, participants are informed of a symbol, either a circle or a triangle, and an identification number, from 1 to 15, assigned to them. Groups are composed of 8 circles (the majority group) and 7 triangles (the minority group). Figure 3a presents the screen that participants see at the start of the experiment in any of the treatments with endogenous networks (note that the positions of circles and triangles are mixed to avoid potential visual biases). For the case of exogenous networks, participants' first screen already portrays the links between nodes (see Figure 3b). In either case, every participant knows his symbol, number and the symbol and number of the 14 others in his group. *Both symbol and number are kept fixed for the entire 20 rounds of play.*

3.2 Treatments with endogenous and exogenous networks

In what follows, we introduce the two main treatments in our study: ENDO and EXO. The main purpose of these treatments is to investigate the role of endogenous linking on the emerging outcome.

Treatment ENDO refers to the situation in which networks are formed endogenously. But the cost of links is zero. There are two stages: first, players simultaneously make linking proposals to any of the other 14 in their group. Reciprocated proposals lead to the creation of links. In the second stage, players are informed of the links proposed and



a. Network formation **b. Action choice**

c. Results **d. Network formation in round** > 1

Figure 3. Screens in the experiment.

Note: (a) Stage 1 in round 1: choosing proposals. The network display illustrates the type and identity number of each player in a group. The decision maker's identity marker is also presented at the bottom of the screen. (b) Stage 2 in any round: choosing *up* or *down*. In the network display, unreciprocated proposals are represented as light 'incomplete' edges, whereas reciprocated proposals are represented as dark 'complete' edges. The decision maker's proposals are highlighted in **red**. (c) End of any round. The decision maker observes own links, and which player(s) chose the same action (in **red**). Payoffs of the decision maker are summarized on the right hand side of the screen. (d) Stage 1 in round $r > 1$: choosing proposals. The decision maker observes a summary of proposals, links, and actions from round $r - 1$.

formed in stage 1. After observing the created network, players choose one of two actions: *up* or *down*. Figure 3b illustrates information about the network that players observe, at this point; in this picture, reciprocated (bilateral) proposals are represented as dark and 'complete' links, while proposals that are not reciprocated are represented as light shorter 'incomplete' edges.[9] Reciprocated and unreciprocated links involving the decision maker are highlighted in red, while any other link is depicted in grey. For example, in the screenshot in Figure 3b, player 14 has links with 2, 7, 8, 9, 10, 11, 12 and 13. He does not reciprocate proposals from 5 and 6, while he makes unreciprocated proposals to 1, 4 and 15. The reciprocated links lead to the relation of being neighbours.

The values of the key parameters in ENDO are as follows:

- $\alpha = 4$: payoff for coordinating on one's *most* preferred action,
- $\beta = 2$: payoff for coordinating on one's *less* preferred action,
- $k = 0$: cost of any bilateral link.

For a player with symbol circle (triangle), his preferred action is *up* (*down*). Every player sees the outcome of the game and his net payoffs on the screen, as in Figure 3c. The figure shows that player 14's neighborhood includes 2, 7, 8, 9, 10, 11, 12 and 13. He coordinates successfully on his preferred action with players 7, 8, 9, 10, 11, 12 and 13, and he fails to coordinate with 2. Thus his net payoff is $8 \times 4 = 32$. Finally, at the beginning of any subsequent round, in stage 1, every player receives information about every other player's links and actions in the previous round, as shown through Figure 3d.

Propositions 2 and 3 tell us that the complete network with conformity on the majority's preferred action is an equilibrium and also socially efficient: see Table 1. Given our parameters, every individual in the minority group earns more by conforming to the majority's preferred action than by choosing his preferred action. So it seems reasonable to postulate the following hypothesis.

[9] An edge departing from node i towards node j without connecting j means that player i proposes a link to player j but j does not propose a link to i

Hypothesis 1. *In the* ENDO *treatment, the outcome will be the complete network and conformity on the majority's preferred action.*

The second treatment is EXO, which differs from ENDO in that there is no network formation stage. Instead, a complete network of interaction is exogenously imposed, so that all players interact with every member of the group. After observing the complete network, as in ENDO, players choose one of two actions: *up* or *down*.[10] Given that there is no linking decision, there are also no linking costs. For this reason, and to make payoffs comparable to ENDO, the parameters in EXO are set as follows:

- $\alpha = 4$: payoff for coordinating on one's *most* preferred action,
- $\beta = 2$: payoff for coordinating on one's *less* preferred action

The detailed instructions handed out to subjects in both treatments, ENDO and EXO, are presented in Appendix E.

In this treatment, the efficient equilibrium involves *conformity* on the majority's preferred action, *up* (see Proposition 3). Moreover, given our parameters, a minority individual actually earns more in this conformity outcome than he would in the diversity outcome where different preference types abide by their preferred actions. We conjecture that behaviour in the game will be indistinguishable from ENDO where subjects coordinate on the socially efficient outcome. Our second hypothesis is:

Hypothesis 2. *In the treatment* EXO*, the outcome will be conformity on the majority's preferred action.*

3.3 Treatments with varying cost of linking

The first two treatments are designed to explore how linking affects coordination and efficiency. For that purpose, we deliberately isolated confounding factors such as the costs associated with establishing a link between two nodes. Yet, the dynamics of link formation

[10] The complete network is shown as it would be in ENDO, had the complete network emerged. See the instructions in Appendix E.

in ENDO may realistically generate network structures that differ, even slightly, from the complete network. For example, subjects may choose to not propose links with players who did not previously propose a link with them, or select their preferred action (e.g., in the previous period). Alternatively, subjects may choose to use their linking activity in the first stage to signal their intended action in the second stage (e.g., disconnecting from players with a different type signals one's intent to not select their preferred action). As a result of generating incomplete networks, such scenarios can create coordination problems in the second stage that are distinct from that in EXO. As a means to better understand the role that linking plays in resolving the coordination problem in the second stage, we manipulate the cost associated with linking in two new treatments.

Treatment SUBSIDY is a treatment with endogenous network formation, and it is different from ENDO in that there is a *small negative* cost for forming links. We set the parameters as

- $\alpha = 4$: payoffs from coordinating on preferred action.
- $\beta = 2$: payoff from coordinating on less preferred action.
- $k = -0.3$: negative cost of a link.

These parameter values lead to a departure from the net payoffs in the treatment ENDO. They have been chosen so as to ensure that conformity on majority's preferred action remains the Pareto dominant action. For more details see Appendix A.1. Negative linking costs can allow the absence of links to work as signals of intent. As not forming a link is costly, we conjecture that individuals should tend toward forming all links. By making connectivity more salient, this in turn would allow a direct comparison between the exogenous complete network and the endogenously generated network. We summarize this discussion in the next hypothesis.

Hypothesis 3. *In the treatment* SUBSIDY, *the complete network with conformity on the majority's preferred action will occur more frequently than in the treatment* ENDO. *As a result, social efficiency will be higher in* SUBSIDY *than in* ENDO.

The next treatment is COST: there is a *positive* cost for forming links. The parameter values are set as follows:

- $\alpha = 4.5$: payoffs from coordinating on preferred action.
- $\beta = 2.5$: payoff from coordinating on less preferred action.
- $k = 0.5$: positive cost of a link.

Observe that being connected with a player who plays one's most preferred action is worth $\alpha = 4$ in EXO and $\alpha - k = 4$ in COST and ENDO. Similarly, for the payoffs from the less preferred action, the payoff is 2 in all three treatments. Positive linking costs can allow links to work as signals of intent. For example, for a minority player to form a link with a majority player indicates a willingness to go along and conform with the majority's preferred action. Similarly not forming a link can signal an intention to stick to one's own preferred action. Thus, positive cost of a link may therefore enable a link to serve as a more potent signal as compared to links in the free linking (i.e., ENDO) or the subsidized linking (i.e., SUBSIDY) scenarios. We therefore expect a closer relation between networks and action choice in COST than under ENDO and SUBSIDY. Given the strategic uncertainty in the game, and the potential losses it can generate (by linking with others who do not coordinate in the second stage), we conjecture that the complete network is less likely to emerge than in the treatments ENDO and SUBSIDY.

Hypothesis 4. *In the treatment* COST, *the network will be segregated and players will choose their preferred actions, thus arriving at segregation with diversity. As a result, social efficiency will be lower than in* ENDO.

The equilibrium outcomes associated with these two treatments are presented in the Appendix A.1 and A.2. However, as reported in Table 1, the integration with conformity outcome remains socially efficient in these two treatments, as under treatment ENDO.

The last two columns refer to additional treatments, that are described in Section 5 below.

Table 1. Experimental treatments.

Note: Description of the initial network structure and parameters of the game by treatment, as well as summary of aggregate payoffs (social welfare) in equilibrium for conformity and diversity.

	Endogenous Networks			Exogenous Networks		
	ENDO	SUBSIDY	COST	EXO	EXOSYM	EXOASYM
Connectivity						
Linking costs (k)	0.0	-0.3	0.5	0.0	0.0	0.0
Number of initial links	0.0	0.0	0.0	105.0	97.0	97.0
Coordination						
Preferred action (α)	4.0	4.0	4.5	4.0	4.0	4.0
Non-preferred action (β)	2.0	2.0	2.5	2.0	2.0	2.0
Aggregated payoffs						
Conformity	690.0	711.0	697.5	690.0	654.0	656.0
Diversity	452.0	465.4	459.5	452.0	452.0	452.0

3.4 Procedures

The experiment was conducted in the Laboratory for Research in Experimental and Behavioural Economics (LINEEX) at the University of Valencia. A total of 540 subjects participated in the study (6 groups of 15 participants in each of the 6 treatments). Participants interacted through computer terminals and the experiment was programmed using z-Tree (Fischbacher 2007).

Each session lasted between 90 and 120 minutes. Upon arrival, subjects were randomly seated in the laboratory. At the beginning of the experiment subjects received printed instructions, which were read out loud to guarantee that they all received the same information (see Appendix E).

After reading the instructions, participants played 5 trial rounds to familiarize with the experiment and payoffs in the game. Trial rounds were not paid and groups were rematched at the beginning of the 20 rounds of actual play. At the end of the experiment each subject answered a debriefing questionnaire.

Earnings were calculated as the total sum of all points accumulated across the 20 rounds of play, using the exchange rate of 50 points = 1 euro. On average participants earned 18 euros, including a 5 euro show-up fee. The standard conditions of anonymity and non-deception were implemented in the experiment, and no one participated in more than one session.[11]

4 Experimental results

In this section, we report findings on the effect that network formation has on coordination, equilibrium selection and efficiency, by comparing settings with *endogenous* and *exogenous* networks.

The data in our experiment consists of the decisions made over 20 periods by groups of 15 participants. In each of the 4 main treatments there are 6 groups, resulting in a total of 480 observations at the group level. Throughout the paper, we test the role of network formation, by running random effects GLS regressions, clustering standard errors on groups. We use dummy variables for treatments as the independent variables. We report two-sided p-values in the text and provide all regressions in Appendix B.[12]

Table 2 reports summary statistics of the main variables of analysis by treatment. Recall the first choice in the game is to create network links (in the endogenous formation treatment). There are two classes of links. Links within-types (WT) are those connecting minority players to other minority players (21 possible links) or majority players to other majority players (28 possible links). The second class of links are between-types (BT), which connect minority players to majority players (56 possible links). The table reports the fraction of links formed out of the maximum possible for each case. Conformity reports the fraction of players in the majority (out of 8) and in the minority (out of 7) choosing the action preferred by the majority: action *up*. Finally, efficiency reports the fraction of total points earned in the network divided by the maximum attainable payoff (conformity in the complete network) when the integration with conformity outcome is chosen.

[11] Regarding the demographics, female participants represent 47% of all subjects in ENDO, and 51% in EXO. All participants are undergraduate students, and the average age is 23 years old. Participants' academic backgrounds are in law, finance, business, economics, pedagogics, tourism, and nursing.

[12] We also analyzed the data using Wilcoxon-Mann-Whitney tests and group averages as the unit of observation. The regression' results are consistent with those of the non-parametric tests.

Table 2. Summary statistics across treatments

Note: Average fractions (percentages) for each of the main variables. Standard deviations in parenthesis. There are no standard deviations for treatments with exogenous networks as links are imposed by design.

	Endogenous Networks			Exogenous Networks		
	ENDO	SUBSIDY	COST	EXO	EXOSYM	EXOASYM
WT-links						
Minority	0.98	0.99	0.88	1.00	1.00	1.00
	(0.06)	(0.05)	(0.14)	-	-	-
Majority	0.98	0.99	0.97	1.00	1.00	1.00
	(0.04)	(0.04)	(0.06)	-	-	-
BT-links	0.41	0.47	0.07	1.00	0.44	0.44
	(0.05)	(0.04)	(0.05)	-	-	-
Conformity						
Minority	0.06	0.03	0.07	0.69	0.31	0.52
	(0.10)	(0.06)	(0.10)	(0.37)	(0.39)	(0.39)
Majority	0.97	0.99	0.99	0.99	0.94	0.98
	(0.05)	(0.04)	(0.03)	(0.03)	(0.18)	(0.06)
Efficiency	0.64	0.68	0.62	0.85	0.76	0.80
	(0.26)	(0.02)	(0.04)	(0.15)	(0.12)	(0.13)

4.1 Endogenous versus exogenous links

Consider the network formation stage in treatment ENDO.[13] We observe that networks are highly connected from round 1 onward and the high rates of connectivity continue over time without much variation. Specifically, subjects create roughly 94.5 links (about 10% of total links are missing), individual degree is on average 12.59 (out of 14 possible links) and there are no differences in connectivity between majority and minority players ($p = 0.673$). For such densely connected networks as in ENDO, it is more illuminating to portray the fraction of links missing instead of the fraction of links formed.

Figure 4 depicts the fraction of links missing within-types (WT) for the majority (solid line) and for the minority (dashed line), as well as the fraction of between-types (BT)

[13] We note that the treatments require a group of 15 subjects to play the same game repeatedly (20 times). In principle, therefore, we should also be considering repeated game effects. In our setting, equilibria of the repeated game will include a sequence of the static game equilibrium, and possibly other more complicated patterns of behavior. In the experiments, subjects converge fairly quickly and behave very much in line with a static equilibrium. The key finding is the contrast in outcomes between the exogenous and the endogenous linking setting. As both these treatments involve repeated interactions, repeated game effects are not central to understanding this difference.

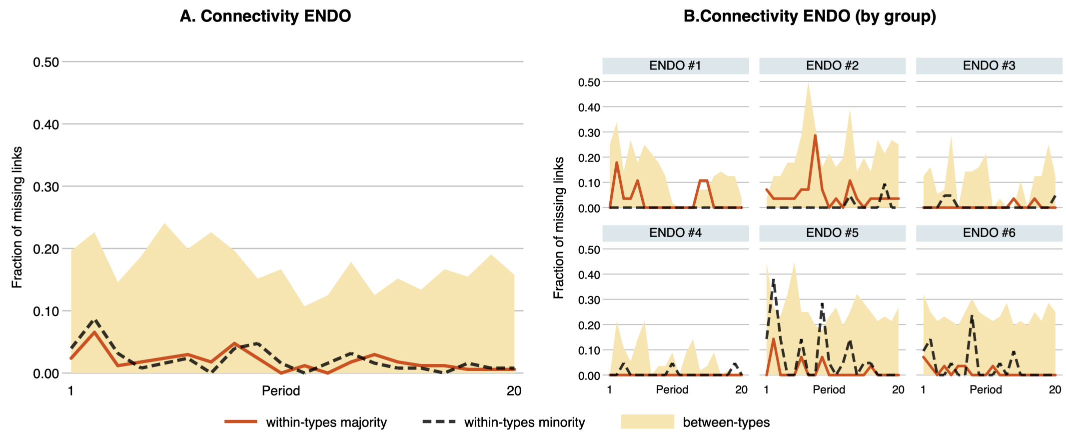


Figure 4. Fraction of missing links in treatment ENDO.

Note: The figure depicts the fraction of between-types missing links (**light area**), as well as the fraction of within-types missing links for the majority (**solid line**) and for the minority (**dashed line**), across the 20 periods of play. Panel A illustrates outcomes pooled at the treatment level and Panel B discriminates by groups.

missing links (light area). We observe that almost all WT-links are formed both for the majority and for the minority, and the fractions are not different between them ($p = 0.872$). Thus, most of the missing connections are BT-links. Moreover, taking a deeper look into the intentions to connect (i.e., proposals made), we find that the likelihood that a proposal is reciprocated and turns into a link is not distinguishable between minority and majority players, neither for WT-proposals ($p = 0.828$) or for BT-proposals ($p = 0.464$). This means that both types of players were equally invested in proposing links and the few missing connections are not caused by a particular type reciprocating less than the other.[14]

We next look at the actions chosen in the coordination game under the two treatments.[15] The main finding is that subjects conform significantly more under treatment EXO than ENDO ($p < 0.0001$): the average number of subjects choosing the majority's action are 12.68 and 8.18, respectively. This is illustrated in Figure 5: the fraction of majority participants

[14]In Appendix D, we report detailed measures on the likelihood of successfully turning proposals into links, by type of player for all endogenous treatments.

[15]Note that conformity on either action is an equilibrium as long as all agents have at least one connection and $\alpha \leq 2\beta$. These conditions are satisfied for all networks endogenously created in ENDO. See details in Table D2 in Appendix D.

(solid line) and minority subjects (dashed line) conforming by choosing the majority's preferred action *up*, across periods. In particular, in five out of the six groups, subjects reached full conformity in treatment EXO (see Figure 5B), while none of the groups in treatment ENDO reaches full conformity even once (see Figure 5D). Thus the experiment rejects hypothesis 1 but strongly supports hypothesis 2.

Furthermore, an inspection of the figure reveals that the main source of the difference between the treatments comes from the choice of the minority, who conform significantly under treatment ENDO ($p < 0.000$); on the other hand, there were no significant differences in the choices of the majority across these two treatments ($p = 0.149$). Finally, note that in treatment ENDO, subjects choose actions in line with their preferences and diversity in actions obtains: this translates into significantly lower level of efficiency as compared to outcomes under treatment EXO ($p < 0.0001$).

These observations constitute the main finding of the paper:

Result 1. ***Endogenous versus exogenous linking***. *In an exogenous complete network, subjects conform to the majority's preferred action; this leads to high levels of efficiency. By contrast, in the endogenous linking game, subjects create an incomplete but dense network, and every subject chooses his preferred action. This leads to diversity in actions and significantly lower level of efficiency than in the exogenous case.*

How is it that subjects form densely connected networks in ENDO, that are very similar to those in EXO, and yet choose diversity in actions? Recall, that the theory predicts that the minority is strictly worse off under diversity as compared to conformity, when the network is complete. The networks that subjects created are dense but not complete: a natural question is whether conformity is still efficient and desirable for the minority subjects in these networks?

To examine this question, we look at the payoffs that would arise if subjects were to conform fully on the majority's action in the created networks and compare these payoffs with the payoffs that subjects actually earn under diversity. We find that efficiency would have been higher had all subjects conformed ($p = 0.000$) to the majority's preferred action in these created networks. This would have especially benefited the majority players who

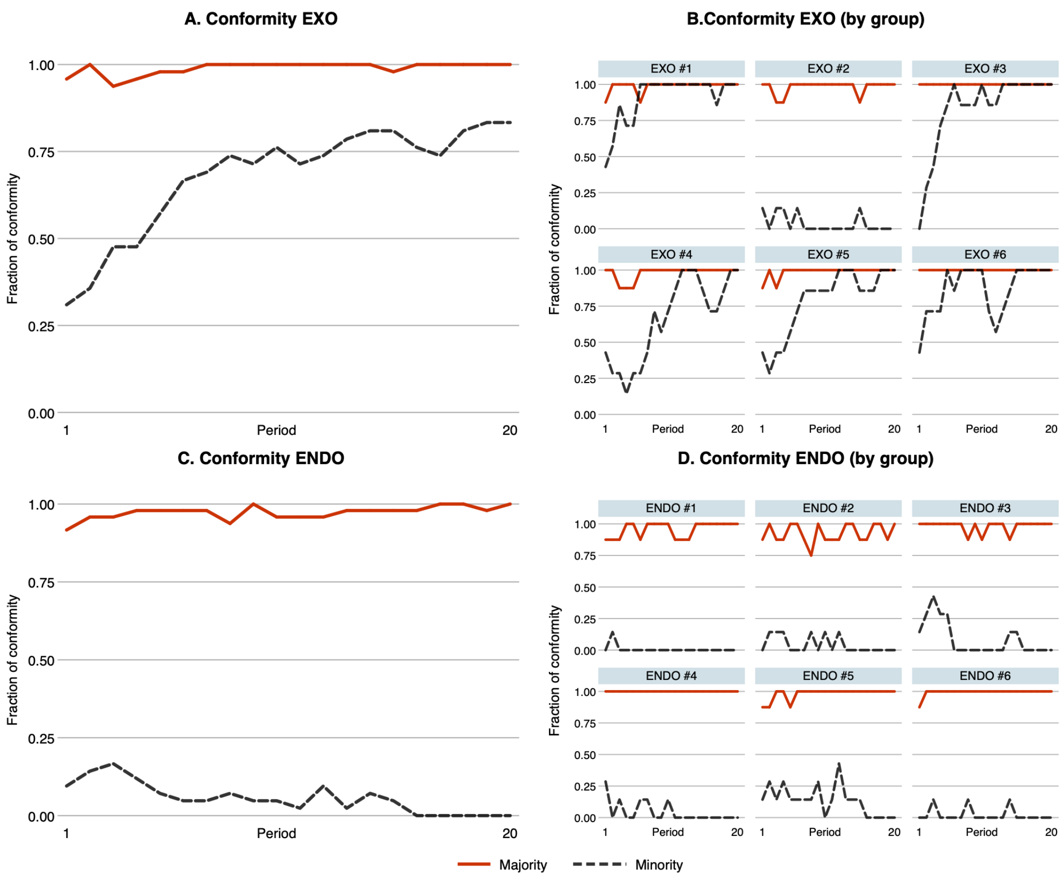


Figure 5. Fraction of subjects choosing conformity (action *up*) in treatments EXO and ENDO.

Note: The figure depicts the fraction of majority (**solid line**) and minority (**dashed line**) subjects choosing action *up*, across periods. Panel A (C) illustrates outcomes pooled at the treatment level and Panel B (D) discriminates by groups for EXO (ENDO).

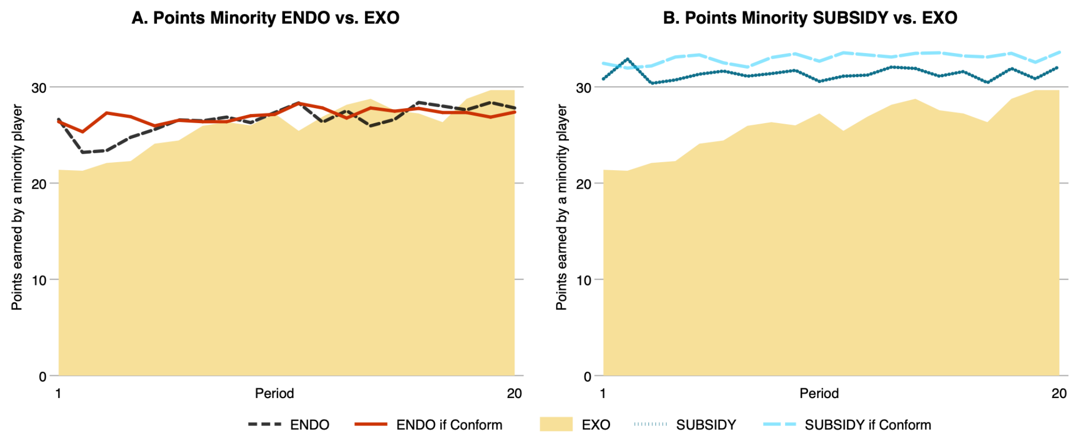


Figure 6. Average payoff for minority players in EXO, ENDO and SUBSIDY.

Note: The **light area** represents the average earnings in EXO in both panels. In Panel A, the **solid line** represents average earnings in ENDO if conformity had been chosen, and the **dashed line** represents the actual average earnings in ENDO. In Panel B, the **long-dashed line** represents average earnings in SUBSIDY if conformity had been chosen, and the **dotted line** represents the actual average earnings in SUBSIDY.

would have earned on average 54 points instead of 31 points. However, minority players would not have seen a significant improvement in earnings: they would have earned on average 27 points instead of 26.3 points (p = 0.647). Figure 6 illustrates the earnings for the minority players under three scenarios: one, under treatment ENDO (dashed line), two, if players had chosen conformity (solid line), and three their earnings under treatment EXO (light area).

This shows us that the minority players are not worse off by choosing diversity compared to choosing conformity, in the incomplete networks that arise under ENDO. We next compare their payoffs under ENDO with their payoffs under treatment EXO. Although efficiency is significantly lower at the network level in ENDO ($p < 0.001$), the actual payoffs attained by the minority in ENDO are not statistically different from what they earn under EXO (p = 0.462). The reason for this, as illustrated in Figure 6, is that coordination on the diversity outcome in ENDO is faster than coordination on the conformity outcome in EXO. The absence of difference in attained payoffs is therefore due to the speed of convergence to conformity under treatment EXO as compared to the rate of convergence to diversity under treatment ENDO. This is best seen if we compare the time trend between treatments in the first and second half of the experiment (i.e., blocks of 10 periods), and find that the time trend is positive and significant for EXO in the first half (p = 0.003) but it is not significant for ENDO (p = 0.221). It is not significant for the second half for either of the two treatments.

These results are consistent with the following view: this is a very complex coordination problem, due to the large number of individuals and the heterogeneity in preferences. Individuals try and use cues from the environment and instruments that they have available to simplify the coordination problem. In our experiment, relatively greater linking with own types correlates strongly with rapid convergence to choosing preferred actions, i.e., to diversity in actions.

We now explore the robustness of this correlation.

4.2 Endogenous incentivized links: the role of negative costs

In this section, we explore more deeply the role of endogenous linking by conducting a treatment with a small negative cost, i.e. a subsidy. In treatment SUBSIDY, any two players can strictly increase their payoffs by 0.3 points when forming a link, regardless of whether they subsequently coordinate their actions. Therefore, we expected this to reduce the fraction of missing links compared to ENDO. If conformity were chosen in the resulting networks, it would suggest that the freedom to choose links does not necessarily lead to different behavior but merely imposes an additional layer of complexity that is hard to solve (thus the missing links), compared to the exogenous case. However, if diversity were chosen, this would provide additional support to our claim that endogenous linking per se affects the way subjects behave in otherwise equivalent social networks.

Figure 7A shows that connectivity is high and that it is higher than under treatment ENDO, 101.3 > 94.5 (p = 0.005). Note that both minority and majority players have high degrees, 13.5 and 13.6, respectively. This is because, on average, both types of players again are indistinguishable in creating all their WT-links (p = 0.538). The few missing connections are BT-links, for which the rate of proposals is also indistinguishable between majority and minority (p = 0.912). This means that both types of players responded

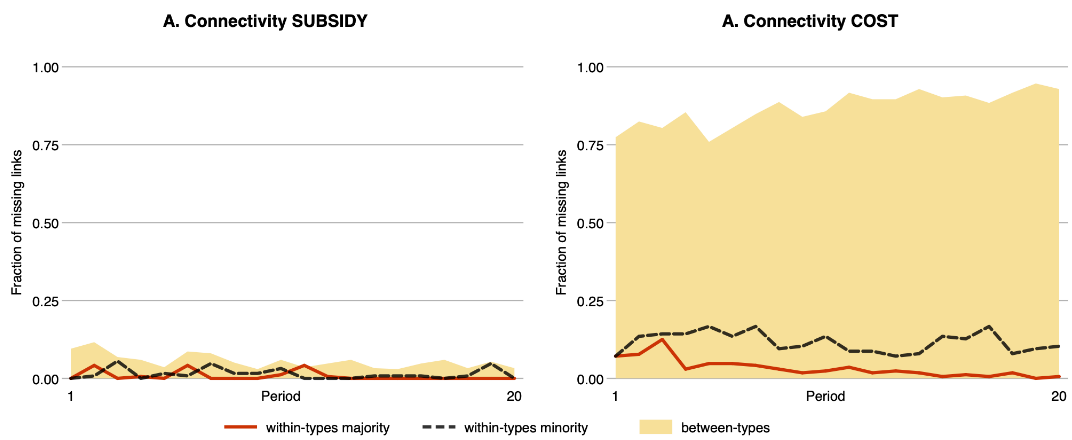


Figure 7. Fraction of missing links in treatment SUBSIDY(Panel A) andCOST (Panel B)

Note: The **light area** represents the fraction of missing links between players with different types. The **solid line** represents the fraction of missing links within players in the majority, and the **dashed line** represents the fraction of missing links within players in the minority.

similarly to the incentives to connect and were highly involved in actively proposing and reciprocating to links.

Turning to action choice in the coordination game, the striking result is that, subjects in SUBSIDY create densely connected networks – in fact they created the complete network in 50% of the cases, and yet not even once did a group reach full conformity. Thus subjects converged to diversity in all the cases. Figure 8A presents patterns of choice in the coordination game. The level of conformity is statistically not distinguishable from ENDO when looking at the choices of the majority ($p = 0.207$) or the minority ($p = 0.108$). The experiment therefore rejects Hypothesis 3.

One point to note is that under treatment SUBSIDY, the network was sufficiently dense so that choosing conformity wuld have actually yielded all subjects strictly higher earnings – so, efficiency would have been higher at the aggregate level ($p = 0.000$), for the majority ($p = 0.000$) as well as for the minority ($p = 0.012$). This suggests that even when the network formation challenge is successfully resolved, the freedom of linking leads to dramatically different behavior in the coordination game.

Result 2. *Negative linking cost.* *When the cost of linking is negative, subjects choose*

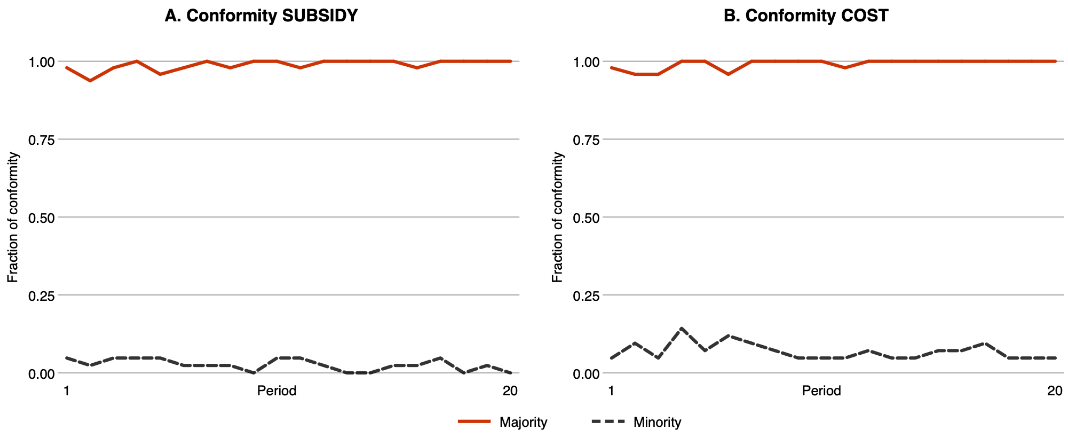


Figure 8. Fraction of subjects choosing conformity (action *up*) in treatments COST and SUBSIDY

Note: The figure depicts the fraction of majority (**solid line**) and minority (**dashed line**) subjects choosing action *up*, across periods.

a dense network – in 50% of the cases, they actually chose the complete network – and yet always also chose diversity of actions.

4.3 Endogenous links: the role of positive costs

To further understand the role of the cost of linking, we also, as a control to the role of linking, conduct a treatment with a positive linking cost. In treatment COST, unlike treatment ENDO, two players should only form a link if they intend to choose the same action in the coordination game. This treatment provides insights into the intentions of majority players. That is, in both ENDO and SUBSIDY, the behavior of the majority players should have been the same if conformity had been chosen, thus the minority was driving outcomes. However, in COST, if majority players want to drive outcomes towards conformity, they should promote BT-links. However, if diversity in actions is chosen, we should see a significant decrease in BT-links compared to ENDO.

Networks in COST are actively created from early on, but this linking activity is mostly focused on WT-links. Thus we see the emergence of (almost) complete segregation in Figure 7B, consistent with Hypothesis 4. Subjects created an average of 53 links out of 105, so that 50% of all possible links are missing. This is justified by the fact that BT-links are very limited at the start, and become rarer over time. The majority forms on average the same number of WT-links in COST and ENDO ($p = 0.259$), while the minority is less successful in forming WT-links when linking is costly than when it is free ($p = 0.029$). However, the main difference is in the BT-links: the rate of missing links increases from 17% in ENDO to 87% in COST ($p < 0.0001$). Moreover, the intention to connect, i.e. fraction of BT-proposals, is not distinguishable between the majority and the minority in COST ($p = 0.775$), which indicates that players in both types are deliberately avoiding connecting between them. Thus, networks converge to two distinct complete components that have virtually no links between them.

We now turn to actions: The main observation is that subjects, in these segregated structures, choose their own preferred action, thereby supporting Hypothesis 4: there is convergence to diversity in actions, and the level of conformity is not distinguishable from ENDO ($p = 0.441$). This is illustrated in Figure 8B.

Result 3. ***Positive linking cost.*** *When the cost of linking is positive, subjects create a network almost completely segregated by types and individuals of the two groups each choose their preferred action, leading to diversity in actions.*

5 Robustness

The previous experimental results suggest a strong effect of choosing links on the selection of the outcome. However, since the endogenously formed networks in the treatment ENDO are not perfectly matching the complete network, there remains the possibility that the existence of a few missing links itself causes the break down of efficient coordination in the second stage.

In order to verify this hypothesis, we provide an alternative examination of the role of endogenous linking compared to exogenous networks. The strategy here is to take dense networks that were created by subjects in the treatment ENDO, set them up as exogenous networks and have the subjects play coordination games on these networks. We take two distinct network configurations with 7 missing links in both, leading to an

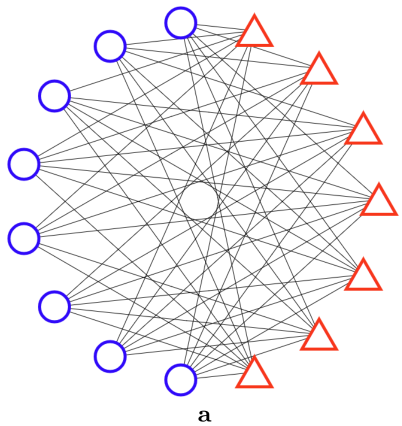


Figure 9. Illustrations of networks in (a) EXOSYM and (b) EXOASYM .

Note: The graphs display the between-types links connecting majority and minority players. For clarity in the illustration, we omit all within-types links, connecting minority to minority players or majority to majority players.

87.5% connectivity across types. We consider one symmetric and one asymmetric pattern of missing links in order to cover extreme cases of the distribution of such missing links. Similarly to the design presented above, each of these new treatments consists of 6 groups of 15 subjects whose decisions are made over 20 periods, resulting in a total of 240 observations at the group level.

Treatment EXOSYM captures a case where the 7 missing links are evenly distributed across the minority players. That is, every minority player has exactly *one* missing link with a majority player (see Figure 9a). After observing the network, as in ENDO, subjects choose one of two actions: *up* or *down*.

Treatment EXOASYM is different from EXOSYM in that the 7 missing links are unevenly distributed. The network is such that only one minority player is missing all but one links with the majority players, while the remaining six minority players are connected to all the majority players (see Figure 9b where the filled triangle node represents the minority player with missing links with all but one majority player).

The interest here is in seeing whether the behavior of subjects remains unchanged or if it is different from the behavior in ENDO. If behavior is markedly different, then that would suggest that the act of linking *per se* is important. The equilibrium characterization

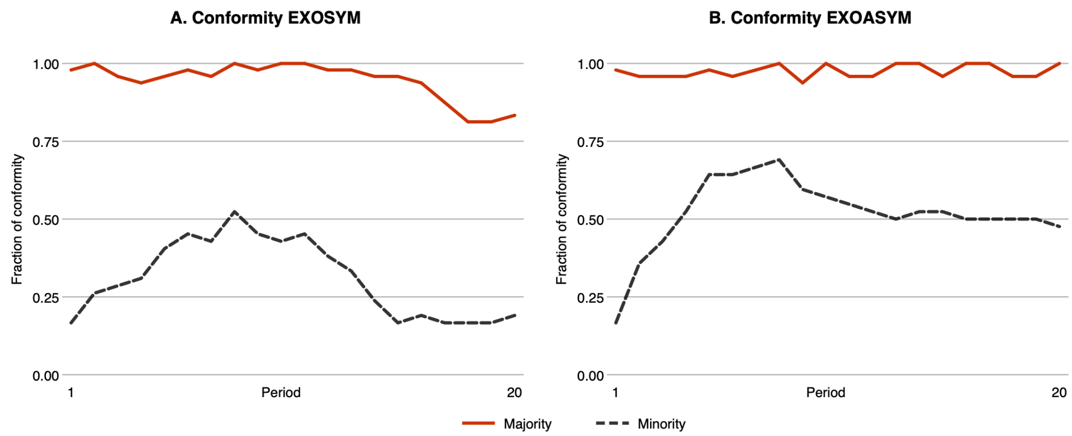


Figure 10. Fraction of subjects choosing conformity (action *up*) in treatments EXOSYM and EXOASYM.

Note: The figure depicts the fraction of majority (**solid line**) and minority (**dashed line**) subjects choosing action *up*, across periods.

for treatments with exogenous-incomplete networks is presented in Appendix 10. Table 1 summarizes the experimental design and the predicted aggregated payoffs (efficiency) in equilibrium.

Our corresponding hypothesis is summarized as follows.

Hypothesis 5. *In the treatments* EXOSYM *and* EXOASYM*, the absence of endogenous linking choices will create significantly different choices of action in the coordination games, as compared to treatment* ENDO*.*

We find that conformity is chosen significantly more by the minority under treatments EXOSYM ($p = 0.05$) and EXOASYM ($p = 0.001$) compared to treatment ENDO. This supports Hypothesis 5. Specifically, under treatment EXOSYM and EXOASYM, four groups (out of twelve) converge to conformity on the majority's preferred action.[16] The diversity outcome was reached in all 100% of the groups under ENDO, but it was attained in only 58% of the groups (7 out of 12) under EXOSYM and EXOASYM. This leads to significantly higher levels of efficiency under EXOSYM ($p = 0.001$) and EXOASYM ($p < 0.001$).

[16] In EXOSYM, one of the remaining groups converges to conformity on the minority's action and the remaining four groups converge to diversity. In EXOASYM, the remaining three groups converge to diversity.

To summarize:

Result 4. ***Exogenous incomplete networks****. When networks are incomplete but exogenously imposed, subjects choose conformity significantly more than in equivalently incomplete but endogenously formed networks. This leads to higher levels of efficiency.*

This sharp difference in outcomes supports the view that the choice of linking *per se* is important in shaping behavior.

6 Possible theoretical explanations

This section examines some of the dominant theoretical approaches to understanding coordination problems that rely upon beliefs and dynamics, and on introspection, respectively. We argue that, despite their relevance to our experimental game, none of them provides an adequate account for the key experimental finding on the differing coordination outcome between the exogenous and the endogenous network treatments.

6.1 Beliefs and dynamics

We start with the approach that focuses on the role of small errors in the process of choice, over time. The idea here is that individuals make small errors or conduct small experiments while dynamically playing the above game and these deviations off the best response help in identifying one of the many (static) equilibrium outcomes. So we will consider a model of dynamics with small perturbations.[17]

First consider an exogenous complete network g. In any round $t > 1$, the dynamic process is described as follows. In each round, a player i is chosen at random to update his strategy x_i^t myopically, best responding to what the other players with whom he interacts did in the previous round, i.e., x_{-i}^{t-1}. There is also a probability $0 < \epsilon < 1$ that a player trembles and chooses a strategy that he did not intend to. Thus, with probability $1 - \epsilon$,

[17]Following the original work of Kandori et al. (1993) and Young (1993), the study of stability in coordination games remains an active field of research; for recent work in this field, see Newton and Angus (2015).

the strategy chosen is $x_i^t = \arg\max_{x_i'} u_i(\theta_i, x_i', x_{-i}^{t-1}, g)$, and with probability ϵ the strategy is $x_i^t \neq \arg\max_{x_i'} u_i(\theta_i, x_i', x_{-i}^{t-1}, g)$. The probabilities of trembles are identical and independent across players, strategies, and rounds. These trembles can be thought of as mistakes made by players or exogenous factors that influence players' choices. Once initial strategies are specified, the above process leads to a well-defined Markov chain where the state is the vector of actions x^t that is played in round t. The Markov chain has a unique stationary distribution, denoted $\mu^\epsilon(x)$. Thus, for any given strategy profile x, $\mu^\epsilon(x)$ describes the probability that x will be the state in some round (arbitrarily) far in the future. Let $\mu = \lim_\epsilon \mu^\epsilon$. According to Young (1993), a given state x is stochastically stable if it is in the support of μ. Thus, a state is stochastically stable if there is a probability bounded away from zero that the system will be in that state according to the steady state distribution, for arbitrarily small probabilities of trembles. In the context of our experiment, Proposition 4 specifies the existence of a unique stochastically stable state in the EXO treatment.

Proposition 4. *Consider an exogenous complete network. If $\frac{\beta}{\alpha} > \frac{n+4}{3n}$, then conformity on the majority's preferred action is the unique stochastically stable outcome.*

The proof is presented in Appendix A.4. According to Proposition 4, stochastic stability provides a clear prediction of convergence to the conformity on the majority's preferred action, which is consistent with our observations from the EXO treatment.

Next consider the endogenous network formation game. Let us simplify the dynamic process by assuming independence of actions in x and linking choices in g such that $X = A^n$ (i.e., as if linking choices and actions were selected simultaneously). Furthermore, let $\overline{g}^t$ denote the network $\overline{g}$ at the end of round t and $s^t = (g^t, x^t)$ denote the action profile at the end of round t (where x^t is as in the exogenous case previously described). In any arbitrary round t, we assume the following dynamic process: (1) first a pair of players ij is randomly picked according to a fixed probability distribution p_{ij} where $p_{ij} > 0$ for each $i, j \in N$. Both players then decide whether to adjust their joint strategies s_{ij} such that it is a best response to s_{-ij}^{t-1} for both i and j (such adjustment may therefore involve adding or severing the link $\overline{g}_{ij}^t$ and/or changing one or both actions x_i^t and x_j^t). Note that $\overline{g}_{ij}^t = 1$ implies that $x_i^t = x_j^t$ even if $x_i^{t-1} \neq x_j^{t-1}$. Similarly, $\overline{g}_{ij}^t = 0$ implies that $x_i^t \neq x_j^t$ even if

$x_i^{t-1} = x_j^{t-1}$. (2) After those choices are made, with probability $0 < \epsilon < 1$, each choice (actions and link) is reversed by a tremble. As a result, there may be up to 3 trembles within a single round t (both actions and the link). This process determines the state s^t according to well-defined probabilities. All trembles and random selections are assumed to be independent in the dynamic process. This leads us to determine stochastic stability across our experimental treatments involving an endogenous network formation.

Proposition 5. *Consider the endogenous linking model where $k \leq 0$. If $\frac{\beta}{\alpha} > \frac{n+4}{3n}$, then integration with conformity on the majority's preferred action is the unique stochastically stable outcome.*

The proof is presented in Appendix A.4. Stochastic stability provides a clear prediction of convergence to integration with conformity on the majority's preferred action whenever $k \leq 0$ (as in Proposition 4). This result is clearly inconsistent with behavior observed in the ENDO and SUBSIDY treatments. Thus, stochastic stability cannot provide an adequate explanation for the behavioral patterns observed in our experiment.

6.2 Team reasoning

Strategic uncertainty is likely to play a major role in explaining people's behavior in our endogenous network formation game. In such a scenario, the obvious difficulty of accurately anticipating every other individual's behavior leads to a search for a 'mechanism' that can be used as a coordination device. Such mechanisms have been studied in the past as possible ways to significantly simplify the framing of the strategic situation from the players' perspective. For example, there is evidence that strategy labeling in games can be effectively used by collectively rational players to coordinate (Sugden 1995; Isoni et al. 2014). Alternatively, it has been argued that situations involving strategic uncertainty can trigger different modes of reasoning. Indeed, as suggested by Bacharach et al. (2006), some individuals may engage in some form of *team reasoning*: they identify themselves as members of a group and conceive that group as a unit of agency acting in pursuit of some collective objective. For example, the collective payoff of a group can be determined as the average individual payoff among its members. In the context of our experiment, a minority (majority) team reasoner would conceive the minority (majority) group as a unit of agency, and as a result would frame the scenario as a two player game between the minority and the majority. This theory assumes that every player of the same type shares the same mental model and consequently acts alike, i.e., for any $i, j \in N$, $x_i = x_j$ if $\theta_i = \theta_j$. This leads us to define a Team Reasoning equilibrium s^* as a strategy profile where no individual $i \in N$ can benefit by a joint deviation of all players of the same type as i.

Formally, for any $i \in N$, $U_i(s^*) = \max_{s_J} U_i(s_J, s^*_{-J})$ where $J = \{j \in J : \theta_j = \theta_i\}$, and $s_J = \prod_{j \in J} x_j$ is a joint strategy of group J.[18] In a complete network, this assumption of same-type similarity in behavior considerably simplifies the decision problem, as summarized in the following result.

Proposition 6. *Consider an exogenous complete network. If $|N_u| > |N_d|$ and $\frac{|N_d|}{n} < \frac{\beta}{\alpha} < \frac{|N_u|}{n}$, then conformity on the majority's preferred action is the unique team reasoning equilibrium.*

The proof is presented in Appendix A.4. Our empirical observations from EXO are consistent with Proposition 6. In particular, the above equilibrium is justified as follows: it is strictly dominant for the majority to play their preferred action, and knowing this, the minority is better off conforming to the majority's preferred action (see the proof of Proposition 6 for details). This difference in the depth of reasoning that is required highlights the difficulty for the minority to reach equilibrium as compared to the majority. The same theory can be applied to the endogenous network formation game, where we similarly assume that every player of the same type shares the same number of proposed links, i.e., for any $i, j \in N$, $x_i = x_j$ and $|g_i| = |g_j|$ if $\theta_i = \theta_j$. This extended assumption of same type similarity in behavior leads to the following result.

Proposition 7. *Let m_u be the number of proposed links by every majority player with the minority ($0 \leq m_u \leq |N_u|$), and m_d be the number of proposed links by every minority*

[18]This equilibrium concept is an extreme case of the unreliable team interaction equilibrium introduced by Bacharach (1999) where all players are assumed to be team reasoners with probability 1.

player with the majority ($0 \le m_d \le |N_d|$)*. If* $|N_u| > |N_d|$, $\frac{|N_d|}{|N_u|} < \frac{\alpha-\beta}{\beta}$, *and* $\frac{|N_d|}{|N_u|-1} \ge \frac{\beta}{\alpha-\beta}$, *then a team reasoning equilibrium is described as one of the following:*

- *Full integration* ($m_u = |N_u|$ *and* $m_d = |N_d|$) *with conformity on the majority's preferred action up.*
- *Segregation* ($m_u = 0$ *and* $m_d = 0$) *with diversity.*
- *Partial integration* ($0 < m_u < |N_u|$ *and/or* $0 < m_d < |N_d|$) *with diversity only if* $k \le 0$.

The proof is presented in Appendix A.4. The conditions in Proposition 7 are consistent with all our experimental treatments involving endogenous linking. In the baseline scenario where $k = 0$, the minority's linking activity in the first stage may play an important signalling role for their subsequent behavior in the second stage. More specifically, all minority players forming links with the majority may signal their joint intention to conform on the majority's preferred action. However, if all minority players propose links with all but one majority players, it would signal their intention to select their preferred action afterwards. While our observations in ENDO are consistent with this kind of equilibrium behavior (according to Figure 4, no more than 90% of links are proposed by the minority to the majority), it is worth noting that this theory alone does not suffice to justify the selection of one particular team reasoning equilibrium, i.e., why did subjects select partial integration with diversity rather than full integration with conformity?

Previous studies (see, e.g., Bardsley et al. (2010); Bardsley and Ule (2017)) have argued that team reasoning is triggered as a means to help people solve complex coordination problems that are too difficult to solve through individualistic reasoning. In our context however, we note that team reasoning does not solve the coordination problem by isolating a unique rational outcome but only reduces the set of available solutions (this multiplicity of equilibrium is highlighted in Proposition 7). As a result, strategic uncertainty remains even among team reasoners, and therefore no clear prediction can be made.

To conclude, although the team reasoning predictions are consistent with the behavior observed in both exogenous and endogenous networks from our experiment, they are insufficient to justify the difference in behavior across those treatments.

6.3 Social preferences

Social preferences have been used to understand behavior in economic settings. Fehr and Schmidt (1999) and Bolton and Ockenfels (2000) argue that people are sensitive to inequality in payoffs and often act to reduce such inequality. One could therefore argue that such inequity aversion can explain results from our experiment. Observe that conformity creates a large gap in payoffs between the minority and the majority, whereas payoffs are relatively similar under heterogeneity. However, this argument applies equally well for the exogenous and for endogenous treatments. But we find that in the treatment EXO, players choose in favor of conformity, while with the same payoff considerations, they choose in favor of diversity in the endogenous linking treatment. If inequity aversion were a strong driving force of behavior, we would expect diversity to emerge in both settings, which is not what we observe.

Alternatively, Charness and Rabin (2002) argue that people may be sensitive to different kinds of social welfare: one may indeed be motivated to help the worst off person ("Maximin" or "Rawlsian" egalitarian criterion) or to maximize the total surplus (classical utilitarianism). In the context of our experiment however, those different motivations lead to aligned preferences (e.g., conformity maximizes both the total surplus and the worst off individual's payoff).

6.4 Bounded reasoning

We next explore the role of limited cognitive abilities. Here we consider cognitive hierarchy theory as introduced by Camerer et al. (2004), according to which players are assumed to be heterogeneous in terms of their depth of reasoning (or reasoning levels). This theory says that naive level 0 players choose at random, level 1 players best respond to expected level 0 players' choices, level 2 players best respond to expected level 1 players' choices, and so on. Applying this theory to the exogenous complete network game from EXO, we obtain the following prediction: as level 0 players will play randomly regardless of their type, level

1 players will best respond by selecting their preferred action (out of 14 other players, 7 are expected to play their preferred action, which is enough according to Proposition 1). If the size of the minority is large enough, as in EXO, then any level m player (with m >1) will best respond to level $m-1$ players by also selecting their preferred action. In other words, diversity is the predicted outcome.

Note that this prediction is robust to the type of naive behavior assumed by the level 0 players. In fact, suppose instead that level 0 players' default behavior is to select their preferred action. In this case, as above, any level m player (m >0) will choose their preferred action as a best response to level $m-1$ players. Our experimental findings under EXO are inconsistent with this prediction.

Hence, established theories of equilibrium selection cannot explain the outcomes we observe in our experiments.

7 Conclusion

This paper studies social coordination in a setting where individuals prefer to coordinate with others but they differ on their preferred action. Our interest is in understanding the role of the choice of linking with others in shaping individual choice.

To clarify the key considerations, we start by setting out a theoretical model. There is a group of individuals who each choose between two actions *up* or *down*. Everyone prefers to coordinate on one action but individuals differ in the action they prefer. We consider a baseline setting in which everyone is obliged to interact with everyone else and a setting in which individuals choose with whom to interact. In the latter setting, everyone observes the network that is created and then chooses between action *up* and *down*. The theoretical analysis reveals a rich set of possibilities.

In the case where everyone interacts with everyone else, there exist three equilibria: everyone conforming to one action, everyone conforming to the other action, and diversity with the two groups choosing their preferred actions. In the setting with endogenous linking the outcomes take two forms: either every individual connects to everyone else and the action profile corresponds to one of the three equilibria described above, or the network is only partially connected. In the latter case, the network may fragment into two components and individuals in each component choose a different action. Finally, we show that in both the exogenous and endogenous interaction setting, conforming to the majority's preferred action maximizes aggregate welfare. Thus there is multiplicity in outcomes both in the exogenous and the endogenous linking case and there is a tension between diversity and aggregate welfare.

Our experiments reveal that, in an exogenous complete network, subjects choose to conform to the majority's preferred action. By contrast, when linking is free and endogenous, subjects form dense networks but choose diverse actions. The networks are biased in favour of linking within same preferences type. An examination of the dynamics of action choice reveals that convergence to the steady state with diverse actions is faster under endogenous linking as compared to the convergence to conformity on the majority's preferred action under the exogenous complete network. Thus our experiments suggest that individuals use links – selectively – to resolve the coordination problems they face.

References

A. Advani and B. Reich. Melting pot or salad bowl: the formation of heterogeneous communities. Institute of Fiscal Studies, WP 15/30, 2015.

A. Antonioni, M. P. Cacault, R. Lalive, and M. Tomassini. Coordination on networks: Does topology matter? PLOS ONE, 2013.

M. Bacharach. Interactive team reasoning: a contribution to the theory of co-operation. Research in economics, 53(2):117–147, 1999.

M. Bacharach, N. Gold, and R. Sugden. Beyond individual choice: teams and frames in game theory. Princeton University Press, 2006.

N. Bardsley and A. Ule. Focal points revisited: Team reasoning, the principle of insufficient reason and cognitive hierarchy theory. Journal of Economic Behavior & Organization, 133:74–86, 2017.

N. Bardsley, J. Mehta, C. Starmer, and R. Sugden. Explaining focal points: Cognitive hierarchy theory versus team reasoning. The Economic Journal, 120(543):40–79, 2010.

M. Bernasconi and M. Galizzi. Coordination in networks formation: Experimental evidence on learning and salience. Coalition Theory Network Working Papers 12159, Fondazione Eni Enrico Mattei (FEEM), 2005.

L. Blume. The statistical mechanics of strategic interaction. Games and Economic Behavior, 4:387–424, 1993.

M. Bojanowski and V. Buskens. Coordination in dynamic social networks under heterogeneity. Journal of Mathematical Sociology, 35:249–286, 2011.

G. E. Bolton and A. Ockenfels. Erc: A theory of equity, reciprocity, and competition. American economic review, pages 166–193, 2000.

C. Camerer. Behavioral game theory: Experiments in strategic interaction. Princeton University Press, 2003.

C. Camerer, T. Ho, and J. Chong. A cognitive hierarchy model of games. Quarterly Journal of Economics, 119(3):861–898, 2004.

G. Charness and M. Rabin. Understanding social preferences with simple tests. The Quarterly Journal of Economics, 117(3):817–869, 2002.

G. Charness, F. Feri, M. A. Melendez-Jimenez, and M. Sutter. Experimental games on networks: Underpinnings of behavior and equilibrium selection. Econometrica, 82:1615–1670, 2014.

S. Choi and J. Lee. Communication, coordination, and networks. Journal of European Economic Association, 12(1):223–247, 2014.

D. Corbae and J. Duffy. Experiments with network formation. Games and Economic Behaviour, 64(1):81–120, 2008.

V. P. Crawford. Adaptive dynamics in coordination games. Econometrica, 63(1):103–143, 1995.

G. Ellison. Learning, local interaction, and coordination. Econometrica, 61:1047–1071, 1993.

L. Ellwardt, P. Hernández, G. Martínez-Canovas, and M. Muñoz-Herrera. Conflict and segregation in networks: An experiment on the interplay between individual preferences and social influence. Dynamic and Games, 3(2):191–216, 2016.

E. Fehr and K. M. Schmidt. A theory of fairness, competition, and cooperation. The quarterly journal of economics, 114(3):817–868, 1999.

U. Fischbacher. z-tree: Zurich toolbox for ready-made economic experiments. Experimental economics, 10(2):171–178, 2007.

J. K. Goeree, A. Riedl, and A. Ule. In search of stars: Network formation among heterogeneous agents. Games and Economic Behavior, 67(2):445–466, 2009.

S. Goyal and F. Vega-Redondo. Network formation and social coordination. Games and Economic Behavior, 50(2):178–207, 2005.

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

M. O. Jackson and A. Watts. On the formation of interaction networks in social coordination games. Games and Economic Behavior, 41(2):265–291, 2002.

M. O. Jackson and A. Wolinsky. A strategic model of social and economic networks. Journal of economic theory, 71(1):44–74, 1996.

M. Kandori, G. Mailath, and R. Rob. Learning, mutation, and long run equilibria in games. Econometrica, 61:29–56, 1993.

M. Kearns, S. Judd, J. Tan, and J. Wortman. Behavioral experiments on biased voting in networks. PNAS, 105(5):1347–1352, 2009.

M. Kearns, S. Judd, and Y. Vorobeychik. Behavioral experiments on a network formation game. ACM EC 2012, 2012.

D. Lewis. Conventions: A Philosophical Study. Harvard University Press, 1969.

F. Lipari, M. Stella, and A. Antonioni. Investigating peer and sorting effects within an adaptive multiplex network model. Games, 10, 2019.

P. R. Neary. Competing conventions. Games and Economic Behavior, 76(1):301–328, 2012.

J. Newton and S. Angus. Coalitions, tipping points and the speed of evolution. Journal of Economic Theory, 157(10):172–187, 2015.

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

T. Schelling. The Strategy of Conflict. Harvard University Press, 1960.

R. Sugden. A theory of focal points. The Economic Journal, pages 533–550, 1995.

H. P. Young. The evolution of conventions. Econometrica, pages 57–84, 1993.

Appendix A Proofs and additional propositions

Proof of Proposition 3:

Proof. Let x and y be the number of players playing *down* in N_u and N_d, respectively. The sum of individual payoffs is

$$W(x,y) = (n - x - y)(\alpha(|N_u| - x) + \beta(|N_d| - y)) + (x + y)(\beta x + \alpha y). \tag{7}$$

For fixed y, social welfare is decreasing in x if $x < x^*$ and increasing in x for $x > x^*$, where

$$x^* = \frac{\beta(|N_d| - 2y) + \alpha(|N_u| - 2y) + \alpha(n)}{2(\alpha + \beta)}. \tag{8}$$

Similarly, for any x, social welfare is decreasing in y if $y < y^*$, and increasing in y for $y > y^*$, where

$$y^* = \frac{\alpha(|N_u| - 2x) + \beta(|N_d| - 2x) + \beta(n)}{2(\alpha + \beta)} \tag{9}$$

Since $0 \leq x \leq |N_u|$ and $0 \leq y \leq |N_d|$, it follows that $W(x,y)$ is maximized for some $x \in \{0, |N_u|\}$ and some $y \in \{0, |N_d|\}$. Note that $W(0, |N_d|) = \alpha(|N_u|^2 + |N_d|^2)$, and $W(|N_u|, 0) = \beta(|N_u|^2 + |N_d|^2)$, which directly implies that $W(0, |N_d|) > W(|N_u|, 0)$ (because $\alpha > \beta$). Furthermore, since $W(0,0) = n(\alpha|N_u| + \beta|N_|)$, we have that $W(0,0) > W(0, |N_d|)$ if and only if

$$\frac{|N_u|}{|N_d|} > \frac{\alpha - \beta}{\alpha + \beta} \tag{10}$$

This inequality holds whenever $|N_u| > |N_d|$.

Similarly, since $W(|N_u|, |N_d|) = n(\beta|N_u| + \alpha|N_d|)$, we have that $W(|N_u|, |N_d|) > W(0, |N_d|)$ if and only if

$$\frac{|N_d|}{|N_u|} > \frac{\alpha - \beta}{\alpha + \beta} \tag{11}$$

This inequality holds whenever $|N_d| > |N_u|$. Furthermore, note that equations (10) and (11) hold for $|N_u| = |N_d|$ as long as $\beta > 0$. To summarize, we always have that either $W(0,0) > W(0,|N_d|)$ or $W(|N_u|,|N_d|) > W(0,|N_d|)$ as long as $|N_u| \neq |N_d|$ or $\beta > 0$.

Finally, consider the case where $x = |N_u|$ and $y = |N_d|$: this implies that $x + y = n$. Since $\alpha > \beta$, it can be shown that $W(0,0) > W(|N_u|,|N_d|)$ so long as $|N_u| > |N_d|$. Moreover, $W(0,0) < W(|N_u|,|N_d|)$ holds as long as $|N_u| < |N_d|$. Finally, $W(0,0) = W(|N_u|,|N_d|)$ if $|N_u| = |N_d|$.

We now show that with endogenous interaction, social welfare is maximized under integration and conformity on the majority's action. The argument is as follows: Start from any network g and any configuration of actions x. Now add all missing links and obtain the complete network. Since $k = 0$ the aggregate payoff remains unchanged. But we know from the first part of the proof that, in the complete network, aggregate payoffs are maximized under conformity on the majority's preferred action. This completes the proof. □

A.1 Treatment SUBSIDY

We start with the case where linking has a negative cost. We obtain the following result.

Proposition 8. *Suppose $k < 0$. Then $(\overline{g}^*, x^*(\overline{g}^*))$ is pairwise stable if one of the following conditions obtain:*

(i) $\overline{g}^*$ *is complete and conformity obtains,* $\forall i \in N$, $x_i^*(\overline{g}^*) = m$, *where* $m \in \{up, down\}$.

(ii) $\overline{g}^*$ *is complete and diversity obtains,* $x_i^*(\overline{g}^*) = \theta_i$ *for all* $i \in N$, *and* $|N_u|, |N_d| \geq \frac{\beta n}{\alpha+\beta} + 1$.

(ii) $\overline{g}^*$ *contains two complete components,* $C_u = N_u$ *and* $C_d = N_d$ *where players in* C_u *choose up, while players in* C_d *choose down, if* $|N_u|, |N_d| \geq \max(\frac{\alpha-\beta}{\beta}, \frac{\beta-k}{\alpha-\beta})$.

The proofs of (i) and (ii) are identical to Proposition 2. The proof of (iii) is however slightly different. The same action profile would still be an equilibrium if only two players of the same component were disconnected, and as a result, such disconnection is not beneficial. However, if two players of different types became connected with each other, the same action profile would still be an equilibrium, but since $k < 0$, both players adding the link would earn strictly more. Conformity on the minority's preferred action would also be an equilibrium in this alternative network (with one link added across types) if $\max(|N_u|, |N_d|) \geq \frac{\alpha-\beta}{\beta}$. In this case, the majority player adding the link would earn $(\max(|N_u|, |N_d|)+1)(\beta-k)$. The payoff for that player in the original segregated network with diversity is $\max(|N_u|, |N_d|)(\alpha - k)$. Therefore, if $\max(|N_u|, |N_d|) \geq \frac{\beta-k}{\alpha-\beta}$, such a player would not benefit from adding a link. A similar argument shows that a minority player cannot benefit from adding a link (because the alternative network may reach conformity on the majority's preferred action). Note that the set of outcomes that are pairwise stable according to Proposition 8 ($k < 0$) are also pairwise stable according to Proposition 2 ($k = 0$). However, the reverse is not true. More specifically, *segregration with diversity* does not always hold for $k < 0$ and specific values for the parameters (see Proposition 8(iii)).

A.2 Treatment COST

Let us now consider the case with costly links. We obtain the following result.

Proposition 9. *Suppose $k > 0$. Then $(\overline{g}^*, x^*(\overline{g}^*))$ is pairwise stable if one of the following outcomes obtains:*

(i) $\overline{g}^*$ is a complete network and conformity obtains, $\forall i \in N$, $x_i^*(\overline{g}^*) = m$, where $m \in \{up, down\}$.

(ii) $\overline{g}^*$ is complete and diversity obtains, $x_i^*(\overline{g}^*) = \theta_i$ for all $i \in N$, and $|N_u|, |N_d| \geq \frac{\beta(n-1)+k}{\alpha}$.

(iii) $\overline{g}^*$ contains two complete components, C_u and C_d; every player in C_u chooses up, while every player in C_d chooses down.

The proofs of (i) and (iii) are identical to Proposition 2. The proof of (ii) differs slightly. The same action profile (diversity) would still be an equilibrium if only two players of the same type were disconnected, and as a result, such disconnection is not beneficial. Conformity on the majority's preferred action would be an equilibrium in the alternative network where one link is deleted between players of different types. The payoff for the minority player deleting the link would then be $(n-1)(\beta-k)$. The payoff for that player in the original complete network with diversity is $\min(|N_u|,|N_d|)\alpha - nk$. Therefore, if $\min(|N_u|,|N_d|) \geq \frac{\beta(n-1)+k}{\alpha}$, then this player does not want to delete the link. The same argument can be made for the majority player deleting the link (assuming the alternative network leads to conformity on the minority's preferred action). Note that the set of outcomes that are pairwise stable according to Proposition 9 ($k > 0$) are also pairwise stable according to Proposition 2 ($k = 0$). However, the reverse is not true. More specifically, *integration with diversity* does not hold for $k > 0$ and specific parameters.

A.3 Treatments EXOSYM and EXOASYM

Finally, we present the equilibrium analysis of the coordination game in these treatments.

Proposition 10. *Suppose $|N_u| > |N_d|$. Fix an incomplete network g in which only $|N_d|$ links are missing between minority and majority players, and the degree of any majority player is at least $n-2$. Suppose x^* is a Nash equilibrium. Then the following outcomes are possible:*

(i) conformity on $m \in \{up, down\}$ if $n \geq \alpha/\beta + |N_d|$.

(ii) diversity with every player choosing their preferred action, if $|N_u|, |N_d| \geq \frac{\beta(n+1)}{\alpha+\beta}$.

Proof. Suppose any conformity outcome in (i). Since the number of missing links between minority and majority players is $|N_d|$, any player must have at least a degree $n - |N_d| - 1$ (lowest degree for a minority player missing all $|N_d|$ links). All players who select their preferred action can clearly not improve their payoff through any deviation. However, the payoff for players selecting their least preferred action is at least $(n-|N_d|)\beta$. Any individual deviation from such players instead yields α. As a result, conformity is an equilibrium whenever $(n-|N_d|)\beta \geq \alpha$, which can be rewritten as $n \geq \alpha/\beta + |N_d|$.

Suppose the diversity outcome in (ii). Since the number of missing links between minority and majority players is $|N_d|$ and $|N_u| > |N_d|$, there must exist at least one majority player with a degree $n-1$ (linked with everyone else). There may also be some minority player(s) with a similar degree (e.g., if some other minority player is missing more than one link). It then directly follows that any such player will earn $|N_y|\alpha$ where $y \in \{u, d\}$. Any unilateral deviation however yields $(n - |N_y| + 1)\beta$. As a result, such a player is not better off deviating if $|N_y|\alpha \geq (n - |N_y| + 1)\beta$, which can be rewritten as $|N_y| \geq \frac{\beta(n+1)}{\alpha+\beta}$. Since other players can only be less connected with the opposite type, they can also not benefit by deviating under this condition. Thus, diversity is an equilibrium. □

The main point to note is that conformity (on *up* or *down*) and diversity both remain equilibrium outcomes under EXOSYM and EXOASYM.

A.4 Beliefs and Dynamics

Proof of Proposition 4:

Proof. Let N_{maj} be the majority group whose members prefer action $x \in \{up, down\}$, i.e., $N_{maj} = \{i \in N : \theta_i = x\}$, and $N_{min} = N \backslash N_{maj}$ represents the minority group in N. The set of absorbing states is characterized by the set of Nash equilibria in pure strategies as specified by Proposition 1. Without loss of generality, let C_{maj} define the conformity outcome where everyone selects the majority's preferred action x ($\in \{up, down\}$), C_{min} define the conformity outcome where everyone selects the minority's preferred action $y \neq x$, and D define the diversity outcome where everyone plays their preferred action. As a result, there are at most three recurrent communication classes each of which corresponds to a particular absorbing state: C_{maj}, C_{min}, and D. We want to determine the resistance of every path between every two recurrent classes, which corresponds to the number of trembles necessary to move from one absorbing state to another. For example, $r(C_{maj}, D)$ determines the resistance from state C_{maj} to state D. According to Proposition 1, every player in the complete network selects their preferred action if m other players in N also select it, such that $\frac{n\beta-\alpha}{\alpha+\beta} < m \leq \frac{n\beta-\alpha}{\alpha+\beta} + 1$. From C_{maj}, it therefore takes at least m

players from N_{min} to switch their action through trembles before it is a best response for the remaining players to switch theirs. As a result, we have $r(C_{maj}, D) = m$. A similar argument leads to $r(C_{min}, D) = m$. From D, it takes at least $|N_{min}| - m$ players from N_{min} to tremble before it is a best response for the remaining players from N_{min} to switch theirs. Therefore, we have $r(D, C_{maj}) = |N_{min}| - m$. A similar argument leads to $r(D, C_{min}) = |N_{maj}| - m$ as it takes $|N_{maj}| - m$ players from N_{maj} to tremble before it is a best response for the remaining players from N_{maj} to switch theirs.

Finally, it is easy to see that $r(C_{maj}, C_{min}) = r(C_{maj}, D) + r(D, C_{min}) = |N_{maj}|$ and $r(C_{min}, C_{maj}) = r(C_{min}, D) + r(D, C_{maj}) = |N_{min}|$.

According to Young (1993), given any state x, an x-tree is a directed graph with a vertex for each state and a unique directed path leading from each state y ($\neq x$) to x. The resistance of x, noted $r(x)$, is then defined by finding an x-tree that minimizes the summed resistance over directed edges. From the above, it is easy to show that $r(C_{maj}) = r(C_{min}, D) + r(D, C_{maj}) = |N_{min}|$, $r(C_{min}) = r(C_{maj}, D) + r(D, C_{min}) = |N_{maj}|$, and $r(D) = r(C_{min}, D) + r(C_{maj}, D) = 2m$. Since $|N_{min}| < |N_{maj}|$, we have $r(C_{maj}) < r(C_{min})$. Moreover, $\frac{\beta}{\alpha} > \frac{n+4}{3n}$ implies that $|N_{min}| < \frac{n}{2} < 2m$, and therefore $r(C_{maj}) < r(D)$. It follows that C_{maj} is the only stochastically stable outcome (Young 1993). □

Proof of Proposition 5:

Proof. Let us first determine the set of absorbing states. It is easy to see that any two players who play the same action must be linked with each other. This implies that the network corresponds to a set of isolated complete components. Moreover, since $|A| = 2$, there can be at most 2 such components. We will refer to any complete network as an integration outcome, and any network with 2 distinct components a segregation outcome. First, it is straightforward to see that any integration outcome with conformity on the same action from A is always stable. Regarding the segregation outcomes, since $k \leq 0$, it is then easy to show that they all belong to the same absorbing state, which consists of the complete network where every player selects their preferred action. In fact, in any such segregation outcome, it is (weakly) dominant for everyone to form links with everyone else.

In the resulting complete network, the only stable diversity outcome is one where every player chooses their preferred action (see proof of Proposition 4 for details).

Regarding the recurrent communication classes, we therefore denote C_{maj} as the integration state with conformity on the majority's action, C_{min} as the integration state with conformity on the minority's action, and D as the integration state with diversity. The proof of Proposition 5 then directly follows from the proof of Proposition 4. □

Proof of Proposition 6:

Proof. Since players of the same type choose the same action, they each earn the same payoff. We then refer to the majority and the minority as single entities. Note that the majority would obtain at least $|N_u|\alpha$ for playing *up*, and at most $n\beta$ for playing *down*. If $\frac{\beta}{\alpha} < \frac{|N_u|}{n}$, then it is strictly dominant for the majority to play *up*. Moreover, the minority would then obtain $|N_d|\alpha$ for selecting *down*, and $n\beta$ for selecting *up* (assuming the majority plays *up*). Since $\frac{\beta}{\alpha} > \frac{|N_d|}{n}$, the minority is then strictly better off selecting *up*. This yields conformity on the majority's action as the only equilibrium solution. □

Proof of Proposition 7:

Proof. We again refer to the majority and the minority as single entities. It is straightforward to see that segregation with diversity is a subgame perfect equilibrium for any $k \geq 0$. Now let us assume that $k \leq 0$. If the majority proposes m_d links with the minority, then playing *down* would at most yield $|N_u|(\beta - k) + k - m_d k$ if the minority plays *up*, and $(N_u + m_d)(\beta - k) + k$ if the minority plays *down*. Similarly, playing *up* would at most yield $(N_u + m_d)(\alpha - k) + k$ if the minority plays *up*, and $N_u(\alpha - k) + k - m_d k$ if the minority plays *down*. Since $\frac{m_d}{|N_u|} \leq \frac{|N_d|}{|N_u|} < \frac{\alpha-\beta}{\beta}$, playing *down* is strictly dominated by playing *up* for the majority, regardless of the links proposed and formed with the minority. Similarly, assuming the minority forms m_u links with the majority, playing *up* would at most yield $(|N_d| + m_u)(\beta - k) + k$, and playing *down* would at most yield $|N_d|(\alpha - k) + k - m_u k$. Since $\frac{|N_d|}{|N_u|-1} \geq \frac{\beta}{\alpha-\beta}$, it follows that the minority prefers *down* if and only if the network is fully integrated (i.e., $m_u = |N_u|$ and $m_d = |N_d|$). As a result, conformity on *up* is compatible only under full integration. Any partial integration or segregation will lead to a diversity outcome where the majority and the minority play their preferred action. □

Appendix B Regression tables

The data in our experiment consists of the decisions made over 20 periods by groups of 15 subjects. In each of the 6 treatments there are 6 groups, resulting in a total of 720 observations at the group level. The tables below report the results associated to random effects GLS regressions with standard errors clustered on groups.

Table B1. Effect of types on network connectivity in ENDO

Note: GLS regressions with group random effects and standard errors clustered on groups (in parenthesis). The omitted category is the MINORITY type. The dependent variable is the share of formed links by types in column I, the fraction of missing links within types in column II, the fraction of failed proposals within types in column III, and the fraction of failed proposals between types in column IV. ***, **, and * indicate statistical significance at the 0.01, 0.05, and 0.10 levels.

	I	**II**	**III**	**IV**
Majority	0.0119	−0.0022	−0.0017	0.0228
	(0.0282)	(0.0139)	(0.0077)	(0.0312)
Period	0.0026*	−0.0019	−0.0011	−0.0023
	(0.0014)	(0.0015)	(0.0009)	(0.0016)
Majority × Period	−0.0003	0.0005	0.0004	0.0004
	(0.0018)	(0.0016)	(0.0009)	(0.0021)
Constant	0.8917***	0.0219**	0.0121*	0.0735***
	(0.0219)	(0.0114)	(0.0065)	(0.0105)
χ^2	7.63*	9.08**	9.26**	5.08
# Obs.	240	240	240	240

Table B2. Effect of endogenous linking on conformity: ENDO vs. EXO

Note: GLS regressions with group random effects and standard errors clustered on groups (in parenthesis). The omitted category is the treatment EXO. The dependent variable is the level of conformity the network in column I, the level of conformity of the majority in column II, and the level of conformity of the minority in column III. ***, **, and * indicate statistical significance at the 0.01, 0.05, and 0.10 levels.

	I	II	III
ENDO	-0.2931^{***}	-0.0168	-0.6088^{***}
	(0.0610)	(0.0117)	(0.1274)
Period	0.0119^{***}	0.0019^{***}	0.0233^{***}
	(0.0032)	(0.0006)	(0.0067)
ENDO × Period	-0.0139^{***}	0.0003	-0.0300^{***}
	(0.0034)	(0.0010)	(0.0069)
Constant	0.8391^{***}	0.9886^{***}	0.6681^{***}
	(0.0599)	(0.0036)	(0.1259)
χ^2	29.50^{***}	18.15^{***}	39.77^{***}
# Obs.	240	240	240

Table B3. Effect of endogenous linking on efficiency: ENDO vs. EXO

Note: GLS regressions with group random effects and standard errors clustered on groups (in parenthesis). The omitted category in columns I and II is treatment ENDO-IF, and the dependent variable is the level of efficiency and the aggregate earnings of the minority, respectively. The omitted category in columns III to VI is treatment EXO, and the dependent variable is the level of efficiency in column III, the aggregate earnings of the minority in the first 10 periods in column IV, the aggregate earnings of the minority in the last 10 periods in column V, and the aggregate earnings of the minority in column VI. ***, **, and * indicate statistical significance at the 0.01, 0.05, and 0.10 levels.

	I	II	III	IV	V	VI
ENDO	-0.2698^{***}	-0.4625	-0.2048^{***}	-0.2216	0.8127	0.7328
	(0.0185)	(1.0092)	(0.0415)	(2.0275)	(2.2152)	(0.9968)
Period	0.0022^{**}	-0.0715^{*}	0.0130^{***}	0.7266^{***}	0.3209	0.4098^{***}
	(0.0011)	(0.0392)	(0.0028)	(0.2386)	(0.2178)	(0.0593)
ENDO × Period	-0.0007	0.1154	-0.0115^{***}	-0.4038	-0.2118	-0.2228^{**}
	(0.0013)	(0.0770)	(0.0029)	(0.3299)	(0.2925)	(0.0889)
Constant	0.9067^{***}	26.9689^{***}	0.8417^{***}	27.3342^{***}	26.0825^{***}	25.7737^{***}
	(0.0176)	(0.6117)	(0.0411)	(1.7254)	(1.4532)	(0.5909)
χ^2	215.65^{***}	14.90^{***}	32.76^{***}	19.92^{***}	2.88	58.04^{***}
# Obs.	240	240	240	120	120	240

Table B4. Effect of types on network connectivity in SUBSIDY

Note: GLS regressions with group random effects and standard errors clustered on groups (in parenthesis). The omitted category is the MINORITY type. The dependent variable is the share of formed links by types in column I, the fraction of missing links within types in column II, the fraction of failed proposals within types in column III, and the fraction of failed proposals between types in column IV. ***, **, and * indicate statistical significance at the 0.01, 0.05, and 0.10 levels.

	I	II	III	IV
Majority	0.0062	−0.0062	−0.0029	0.0029
	(0.0203)	(0.0101)	(0.0053)	(0.0271)
Period	0.0017***	−0.0004	−0.0002	−0.0019***
	(0.0005)	(0.0003)	(0.0001)	(0.0006)
Majority × Period	0.0001	−0.0004	−0.0003	0.0013
	(0.0008)	(0.0007)	(0.0003)	(0.0008)
Constant	0.9614***	0.0141	0.0074	0.0267**
	(0.0166)	(0.0097)	(0.0051)	(0.0145)
χ^2	17.93***	5.15	5.23	14.26***
# Obs.	240	240	240	240

Table B5. Effect of negative linking cost on network connectivity: SUBSIDY vs. ENDO

Note: GLS regressions with group random effects and standard errors clustered on groups (in parenthesis). The omitted category is the treatment ENDO. The dependent variable is the fraction of missing links within types for the majority in column I, and for the minority in column II, the fraction of failed proposals within types for the majority in column III, and for the minority in column IV, the fraction of failed proposals between types for the majority in column V, and for the minority in column VI. ***, **, and * indicate statistical significance at the 0.01, 0.05, and 0.10 levels.

	I	II	III	IV	V	VI
SUBSIDY	−0.0119	−0.0079	−0.0060	−0.0048	−0.0667*	−0.0469***
	(0.0084)	(0.0149)	(0.0045)	(0.0082)	(0.0373)	(0.0179)
Period	−0.0014***	−0.0019	−0.0008***	−0.0011	−0.0019	−0.0023
	(0.0005)	(0.0015)	(0.0003)	(0.0009)	(0.0013)	(0.0016)
SUBSIDY × Period	0.0006	0.0015	0.0003	0.0009	0.0013	0.0003
	(0.0008)	(0.0015)	(0.0004)	(0.0009)	(0.0015)	(0.0017)
Constant	0.0198**	0.0219*	0.0105**	0.0121*	0.0964***	0.0735***
	(0.0079)	(0.0114)	(0.0042)	(0.0065)	(0.0294)	(0.0105)
χ^2	9.61**	7.70*	9.71**	6.76*	5.85	34.99***
# Obs.	240	240	240	240	240	240

Table B6. Effect of negative linking cost on conformity: SUBSIDY vs. ENDO

Note: GLS regressions with group random effects and standard errors clustered on groups (in parenthesis). The omitted category is the treatment ENDO. The dependent variable is the level of conformity of the majority in column I, the level of conformity of the minority in column II, the level of conformity in the network in column III, and the total earnings of the minority players in column IV. ***, **, and * indicate statistical significance at the 0.01, 0.05, and 0.10 levels.

	I	II	III
SUBSIDY	0.0159	−0.0324	−0.0066
	(0.0126)	(0.0201)	(0.0122)
Period	0.0022***	−0.0068***	−0.0019*
	(0.0008)	(0.0019)	(0.0012)
SUBSIDY × Period	−0.0005	0.0052**	0.0022
	(0.0011)	(0.0023)	(0.0014)
Constant	0.9718***	0.0593***	0.5459***
	(0.0111)	(0.0193)	(0.117)
χ^2	14.60***	18.58***	3.81
# Obs.	240	240	240

Table B7. Effect of endogenous linking on efficiency: SUBSIDY vs. EXO

Note: GLS regressions with group random effects and standard errors clustered on groups (in parenthesis). The omitted category in columns I and II is treatment SUBSIDY-IF, and the dependent variable is the level of efficiency in column I, and the aggregate earnings of the minority in column II. The omitted category in columns III and IV is treatment EXO, and the dependent variable is the level of efficiency in column III, and the aggregate earnings of the minority in column IV. ***, **, and * indicate statistical significance at the 0.01, 0.05, and 0.10 levels.

	I	II	III	IV
SUBSIDY	−0.2929***	−1.6173**	−0.1669***	5.5667***
	(0.0126)	(0.6401)	(0.0412)	(0.6881)
Period	0.0016***	0.0532***	0.0129***	0.4098***
	(0.0006)	(0.0155)	(0.0028)	(0.0593)
SUBSIDY × Period	−0.0010*	−0.0417	−0.0124***	−0.3983***
	(0.0006)	(0.0591)	(0.0028)	(0.0823)
Constant	0.9677***	32.9577***	0.8417***	25.7727***
	(0.0124)	(0.5341)	(0.0411)	(0.5909)
χ^2	540.83***	17.52***	30.93***	86.25***
# Obs.	240	240	240	240

Table B8. Effect of positive linking cost on network connectivity: COST vs. ENDO

Note: GLS regressions with group random effects and standard errors clustered on groups (in parenthesis). The omitted category is the treatment ENDO. The dependent variable is the fraction of missing links within types for the majority in column I, and for the minority in column II, the fraction of failed proposals within types for the majority in column III, and for the minority in column IV, the fraction of failed proposals between types for the majority in column V, and for the minority in column VI. ***, **, and * indicate statistical significance at the 0.01, 0.05, and 0.10 levels.

	I	II	III	IV	V	VI
COST	0.0150	0.0950**	0.0076	0.0461**	0.4825***	0.4573***
	(0.0133)	(0.0434)	(0.0069)	(0.0209)	(0.0895)	(0.0767)
Period	−0.0014***	−0.0019	−0.0008***	−0.0011	−0.0019	−0.0023
	(0.0005)	(0.0015)	(0.0003)	(0.0009)	(0.0013)	(0.0016)
COST × Period	−0.0027	0.0004	−0.0014	0.0005	0.0086***	0.0094
	(0.0017)	(0.0035)	(0.0009)	(0.0019)	(0.0028)	(0.0078)
Constant	0.0198**	0.0219*	0.0105**	0.0121*	0.0964***	0.0735***
	(0.0079)	(0.0114)	(0.0042)	(0.0065)	(0.0294)	(0.0105)
χ^2	14.52***	20.64***	14.88***	23.71***	41.49***	71.87***
# Obs.	240	240	240	240	240	240

Table B9. Effect of positive linking cost on conformity: COST vs. ENDO

Note: GLS regressions with group random effects and standard errors clustered on groups (in parenthesis). The omitted category is the treatment ENDO. The dependent variable is the level of conformity of the majority in column I, the level of conformity of the minority in column II, the level of conformity in the network in column III, and the total earnings of the minority players in column IV. ***, **, and * indicate statistical significance at the 0.01, 0.05, and 0.10 levels.

	I	II	III
COST	0.0191*	0.0105	0.0151
	(0.0116)	(0.0379)	(0.0196)
Period	0.0022***	−0.0068***	−0.0019*
	(0.0008)	(0.0019)	(0.0012)
COST × Period	−0.0006	0.0051**	0.0021
	(0.0011)	(0.0023)	(0.0012)
Constant	0.9718***	0.0593***	0.5459***
	(0.0111)	(0.0193)	(0.0117)
χ^2	12.93***	25.99***	6.21
# Obs.	240	240	240

Table B10. Effect of incomplete (symmetric) exogenously-fixed networks on conformity: EXOSYM vs. ENDO

Note: GLS regressions with group random effects and standard errors clustered on groups (in parenthesis). The omitted category is the treatment ENDO. The dependent variable is the level of conformity of the majority in column I, the level of conformity of the minority in column II, the level of conformity in the network in column III. ***, **, and * indicate statistical significance at the 0.01, 0.05, and 0.10 levels.

	I	II	III
EXOSYM	−0.0235	0.2536**	0.1058*
	(0.0311)	(0.1291)	(0.0622)
Period	0.0022***	−0.0069***	−0.0019*
	(0.0008)	(0.0019)	(0.0012)
EXOSYM × Period	−0.0093	−0.0024	−0.0061
	(0.0086)	(0.0103)	(0.0075)
Constant	0.9718***	0.0593***	0.5459***
	(0.0111)	(0.0193)	(0.0117)
χ^2	8.75**	19.81***	8.74**
# Obs.	240	240	240

Table B11. Effect of incomplete (asymmetric) exogenously-fixed networks on conformity: EXOASYM vs. ENDO

Note: GLS regressions with group random effects and standard errors clustered on groups (in parenthesis). The omitted category is the treatment ENDO. The dependent variable is the level of conformity of the majority in column I, the level of conformity of the minority in column II, the level of conformity in the network in column III, and the total earnings of the minority players in column IV. ***, **, and * indicate statistical significance at the 0.01, 0.05, and 0.10 levels.

	I	II	III	IV
EXOASYM	0.0027	0.4581***	0.2152***	
	(0.0159)	(0.1343)	(0.0688)	
Period	0.0022***	−0.0068***	−0.0019*	
	(0.0008)	(0.0019)	(0.0012)	
EXOASYM × Period	−0.0012	0.0099	0.0040	
	(0.0016)	(0.0112)	(0.0054)	
Constant	0.9718***	0.0593***	0.5459***	
	(0.0111)	(0.0193)	(0.0117)	
χ^2	7.76*	38.35***	18.54***	
# Obs.	240	240	240	

Appendix C Additional Experiments

C.1 Different costs of linking

This section presents two additional experiments we ran: SUBSIDY- and COST+. SUBSIDY- is a treatment with lower subsidy compared to SUBSIDY, given by $k = -0.1 > -0.3$. the game is as in SUBSIDY so that the value of other parameters remain $\alpha = 4$ and $\beta = 2$. COST+ is a treatment with higher linking cost than in the positive cost treatment COST, given by $k = 2 > 0.5$. The game is as in COST, but we set the value of other parameters at $\alpha = 6$, $\beta = 4$.

First, we look at the main findings in the SUBSIDY- treatment. Consistent with the results from SUBSIDY, we observe in Figure C1B that a positive subsidy for linking (negative cost) increases the level of connectivity, so that there are no significant differences in network density between SUBSIDY- and SUBSIDY ($p = 0.531$). This is particularly clear when observing Figure C1A. No differences are observed in the WT-types missing links for the majority ($p = 0.865$) or the minority ($p = 0.306$). As in SUBSIDY, most of the missing connections in SUBSIDY- are BT-type links, which also do not differ between the two treaments ($p = 0.674$). Diversity of actions is a prominent outcome in SUBSIDY- with 50% of the groups portraying complete diversity (see Figure C2A)[19]. Thus, a positive subsidy promotes integration and to a large extent diversity in actions.

For the second treatment, COST+, we observe that increasing the cost of linking leads to lower linking across preference types compared to COST ($p = 0.001$). Thus, we observe (almost complete) segregation and diversity (see Figure C1B). Notably, in COST+ individuals choose actions more or less in line with their preferences. The majority choose its preferred action almost from the start and persist with it for the entire experiment as in COST ($p = 0.147$). The minority players mostly choose their preferred action during the first rounds of play and by round 11 no minority player is choosing the action of the majority, which also does not differ compared to COST ($p = 0.154$). The effect of linking

[19] While 3 out of the 6 groups converged to diversity, the 3 remaining groups converged to conformity on the majority's preferred action.

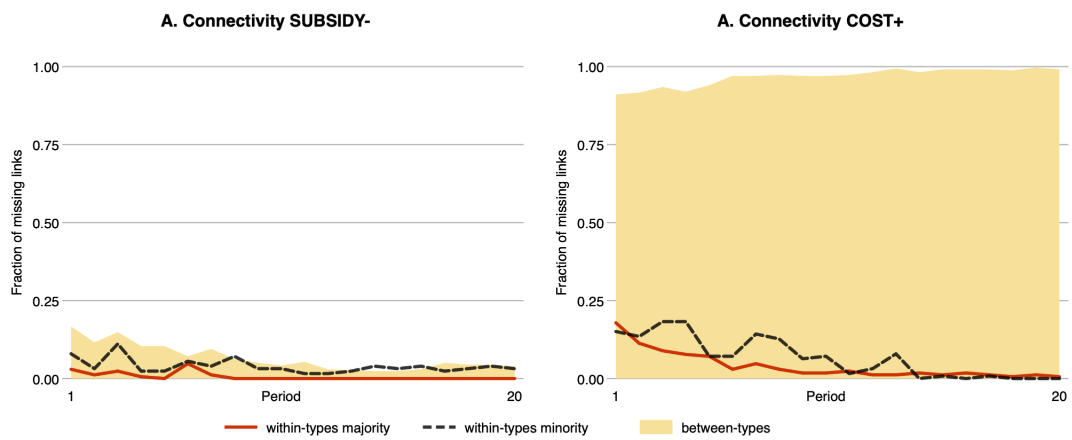


Figure C1. Fraction of missing links in treatments SUBSIDY- and COST+

Note: The **light area** represents the fraction of missing links between players with different types. The **solid line** represents the fraction of missing links within players in the majority, and the **dashed line** represents the fraction of missing links within players in the minority.

costs is clear, segregation and diversity.

C.2 Small minority representation

We briefly studied the role of the size of the minority. As it falls, the payoff losses of separating itself rise. This may induce greater integration and conformity. To test this hypothesis, we conducted an SMALLMIN treatment in which we varied group composition, from 8 majority and 7 minority players (main experimental treatments) to 12 majority and 3 minority players. In this case, if the minority players are excluded by the majority for not conforming, they will be better off seeking redemption than segregating. There were 6 groups in SMALLMIN, the same as in all other treatments. The game and parameters are as in COST+, i.e., with a cost of linking $k = 2$, and we only vary the group composition.

In SMALLMIN, having a larger majority and a smaller minority results in a densely-connected network. Out of the 105-possible links that can be formed, individuals in SMALLMIN form on average 94.61 links across groups and rounds, making networks equivalent to those formed in ENDO ($p = 0.887$). The majority formed significantly more links than the minority across rounds, 12.84>11.73 ($p = 0.000$). As illustrated in Figure C3A, it took the minority longer to link within their type than it took the majority ($p = 0.039$). But more

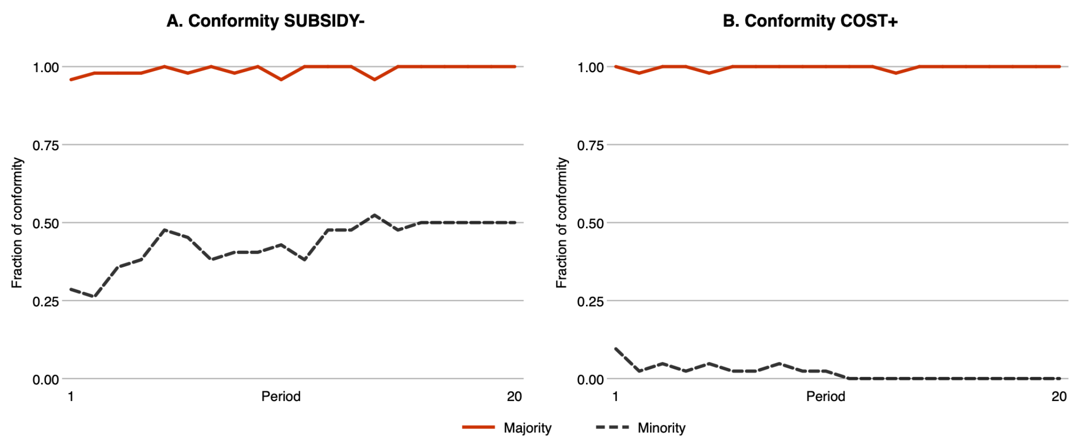


Figure C2. Fraction of subjects choosing conformity in treatments SUBSIDY- and COST+

Note: The **light area** represents the fraction of missing links between players with different types. The **solid line** represents the fraction of missing links within players in the majority, and the **dashed line** represents the fraction of missing links within players in the minority.

importantly, the majority reciprocated significantly less to the proposals of the minority to connect ($p = 0.000$), which only converged after various attempts. On average (across groups and rounds), the minority proposed 10.5 links to the majority in the first five rounds and only 7.6 links were formed. The majority, on the other hand, proposed 2.0 and formed 1.9 links with the minority in the same block. While the majority players were at first reluctant to create links with the minority, the minority players insisted on linking. This persistence appears to have triggered reciprocity in the following rounds.

We observe that the initial reluctance of the majority to form connections with the minority breaks down due to both the persistence of link proposals from the minority and due to the behavior of the minority. More specifically, we observe that from the first round, 78% of the minority players conformed (i.e. chose the action they preferred the least), and by round 3 all minority players were conforming completely. Thus, minority players conform significantly more in SMALLMIN than in ENDO ($p = 0.000$). In fact, once the network is complete, we observe that all players conform to the same action, see Figure C3B. In summary, we conclude that when the minority group size is significantly smaller than the majority group size, subjects converge rapidly to integration and conformity on

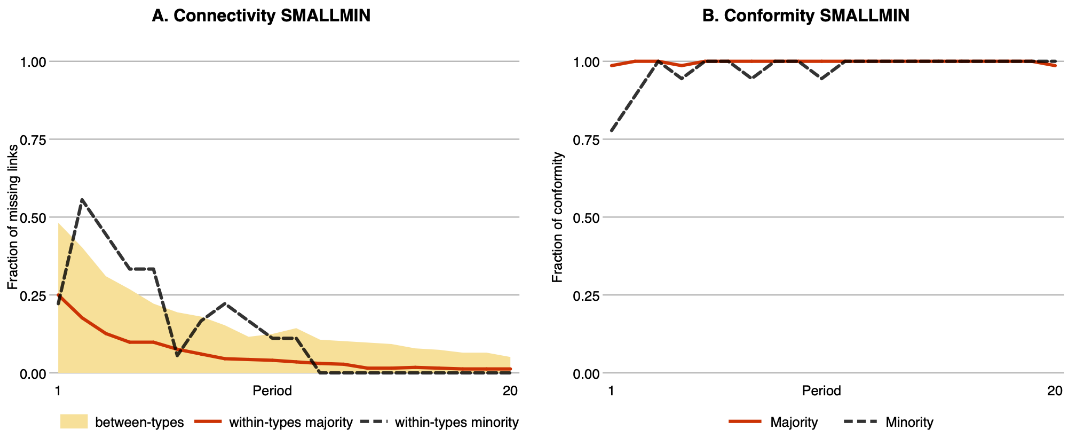


Figure C3. Fraction of missing links (Panel A) and fraction of subjects choosing conformity (Panel B) in treatment SMALLMIN

Note: The **light area** represents the fraction of missing links between players with different types. The **solid line** represents the fraction of missing links within players in the majority, and the **dashed line** represents the fraction of missing links within players in the minority.

the majority's preferred action.

Appendix D Network measures

In this section we report details on the success players have on forming links in their network. The success rates are informative of the ability players have of anticipating reciprocity from others, both within and between types. In Table D1 we report summary statistics on link proposals and links formed at the individual level, discriminated by the type of player making the proposal (majority or minority), the type player being targeted by the proposal (within or between) and the treatment. In all treatments but the SMALLMIN, the majority players can propose/form 7 WT-connections and 7 BT-connections, while the minority can propose/form 6 WT-connections and 8 BT-connections.

Table D1. Summary statistics across treatments

Note: Average number of the main variables. Standard deviations in parenthesis. There are no standard deviations for treatments with exogenous networks given links are imposed by design.

	Main Treatments			Additional Treatments		
	ENDO	SUBSIDY	COST	SUBS-	COST+	SMALLMIN
Minority						
Proposals Within	5.94	5.96	5.59	5.87	5.79	1.83
	(0.42)	(0.39)	(0.99)	(0.64)	(0.86)	(0.49)
Links Within	5.87	5.92	5.30	5.76	5.60	1.73
	(0.49)	(0.43)	(1.24)	(0.71)	(1.02)	(0.58)
Proposals Between	7.15	7.76	2.14	7.74	0.95	11.18
	(2.02)	(1.15)	(3.11)	(1.07)	(2.20)	(1.69)
Links Between	6.63	7.56	1.05	7.47	0.26	10.00
	(1.97)	(1.22)	(1.84)	(1.11)	(0.73)	(2.14)
Majority						
Proposals Within	6.93	6.97	6.88	6.98	6.84	10.60
	(0.39)	(0.40)	(0.61)	(0.30)	(0.59)	(1.18)
Links Within	6.87	6.95	6.77	6.95	6.72	10.34
	(0.48)	(0.43)	(0.69)	(0.33)	(0.70)	(1.37)
Proposals Between	6.41	6.81	2.20	6.71	0.89	2.56
	(1.51)	(0.90)	(2.81)	(1.24)	(1.95)	(0.90)
Links Between	5.80	6.61	0.92	6.54	0.23	2.50
	(1.59)	(1.01)	(1.22)	(1.29)	(0.61)	(0.93)

We observe that successful connectivity is persistent across treatments. This is illustrated in Figure D1. In ENDO minority (majority) players form 99% of the WT-proposals they make to other minority (majority) players. The BT-proposals are also very successful but slightly less than within, 93% and 90% for minority and majority, respectively. Similarly, in SUBSIDY majority and minority are very successful in turning proposals into links, 99% and 97% for WT and BT proposals, respectively, the same for minority and majority players. In COST the WT-proposals are also highly successful, 95% and 98% for minority and majority, respectively. However, the likelihood of turning BT-proposals into links is much lower than in the other treatments: 49% and 41% for minority and majority, respectively. Table D1 and Figure D1 also report connectivity choices for the additional treatments with endogenous networks: COST+, SUBS- and SMALLMIN.

Finally, we explore for each treatment, the instances where players are not successful at all in forming any connections. This is important because conformity on the majority's preferred action is an equilibrium (but not the only one) in any network except one in which a minority player is an isolate. If this were the case, any minority player is better off choosing his preferred action and earning $\alpha > \beta$. However, if a minority player has at least one link, then miscoordinating with his neighbor and choosing his preferred action can never be better than coordinating on his disliked action as long as $2 \times \beta \geq \alpha$. Note that this constraint is satisfied across all our treatments. Table D2 reports the minimum degree as well as the number of isolates for each type and treatment. Notably, in only one of the 120 observations in COST there is an isolate minority player, in period 10. This means that in 99% of the networks formed conformity is an equilibrium in COST and in 100% of ENDO and SUBSIDY. Table D2 also reports

Table D2. Minimum degree count by treatments

Note: The table reports the minimum degree and the count of isolates by type of player in each treatment (out of 120 observations).

	Main Treatments			Additional Treatments		
	ENDO	SUBSIDY	COST	SUBS-	COST+	SMALLMIN
Minority						
Minimum degree	5	1	0	0	1	3
# isolates	0	0	1	4	0	0
Majority						
Minimum degree	5	0	0	0	5	2
# isolates	0	2	1	1	0	0

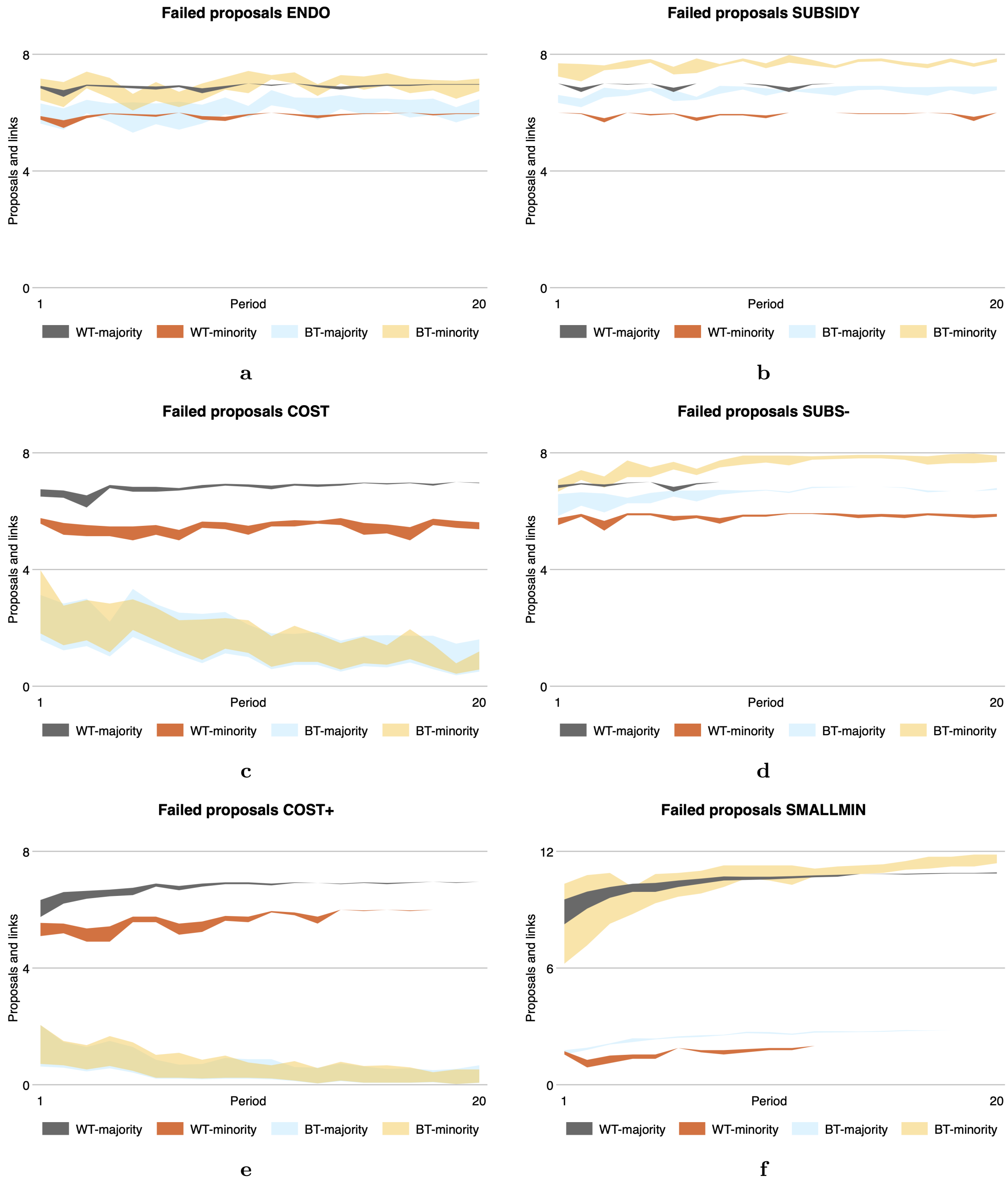


Figure D1. Failed proposals by treatment.

Note: In each range area the upper line is the number of proposals made and the lower line the number of links formed. The inside of an area displays the amount of failed (unreciprocated) proposals by type of player (majority or minority) and type of target (within or between).

Appendix E Instructions

All treatments:

You are participating in an economic experiment where you have to make decisions. For participating in this experiment, you will receive a minimum payment of 5€. Please, read carefully these instructions to find out how you can earn **additional money**.

All interactions between you and the other subjects take place through the computers. Please, do not talk to the other subjects or communicate with them in other way. If you have questions, raise your hand and an experimentalist will come to you to answer it.

This experiment is **anonymous**. Therefore, your identity will not be revealed to the other subjects nor theirs to you.

In this experiment, you can earn points. At the end of the experiment, those points will be converted to Euros using the following exchange rate: 50 points = 1€. You will receive your earnings in cash.

This experiment is composed by 2 identical stages. The first stage is a trial stage, it lasts 5 rounds and the points you earn will not be exchanged for Euros. The second stage is the real experiment, it lasts 20 rounds, and the points you earn will be exchanged for Euros at the end of the experiment. Next, you will be informed of the decisions to you can make in each round.

Decisions in each round

At the beginning of each round, all subjects are randomly assigned to groups of size 15. You will be in a group with the same people for an entire stage. Please, remember that the first stage is a trial stage (5 rounds), and the second is the experiment (20 rounds).

Each subject in a group is randomly assigned a symbol (**circle or triangle**) and a number (**between 1 and 15**). You will be informed about your number and your symbol at the bottom of your screen, which will not change within a stage. That is, your number and your symbol might change from the trial stage to the experiment stage, but not between the rounds of a given stage.

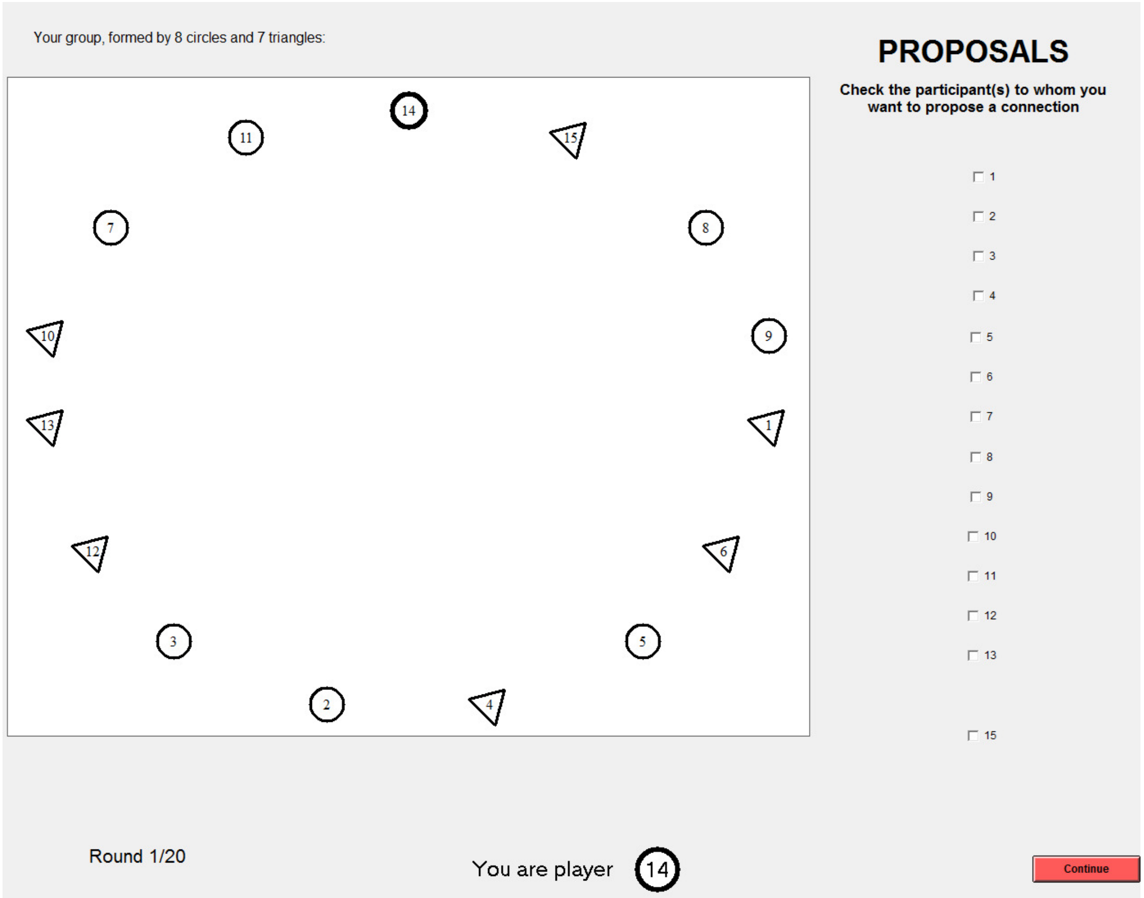
Specific to Treatment ENDO only:

Each round consists of 3 phases: (1) Linking, (2) Action and (3) Earnings.

Phase 1. Linking

At the beginning of the first round you will see the interaction network formed in the previous round. Naturally, in round 1 you will see an empty network. You will see your number and your type, and the numbers and types of the other subjects, as illustrated in the image below. You will be highlighted with a thicker border, to facilitate that you can identify yourself in the screen.

The first decision you make regards whom you want to propose a connection to. You can propose between 1 and 14 connections. To do so, you have to click the checkbox next to a subject's number, in the list on the right hand side of the screen. Once you checked all the proposals you want to make, click the Continue button.



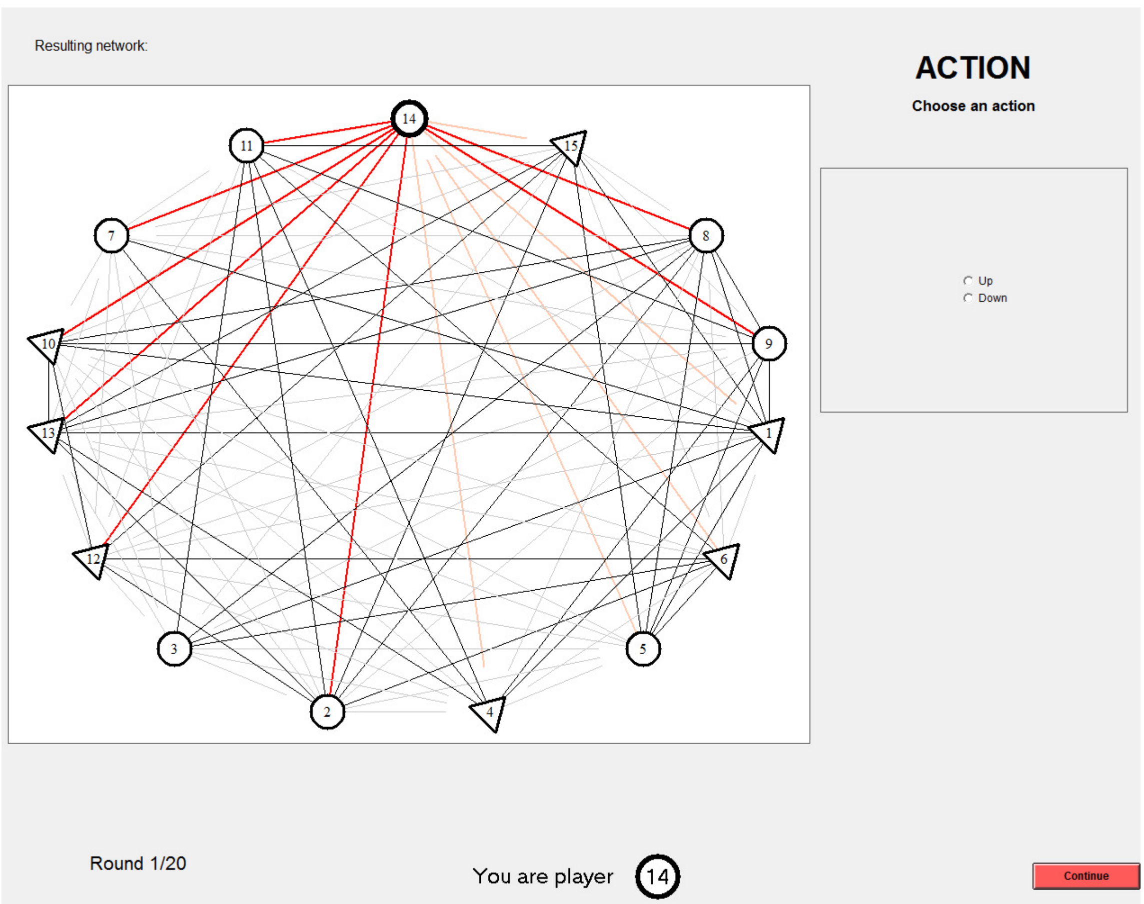
A connection is formed if 2 subjects propose to each other. In Phase 2 (Action) you will interact only with the subjects to whom are connected.

Phase 2. Action

Once all subjects have made all their proposals, you will see the resulting network of interactions. A line starting from you and reaching another subject represents a connection between you and the other subject. A thinner line starting from you, directed to another subject, without reaching him, represents a proposal you made to the such subject, which he did not reciprocate. Similarly, a line starting from other subject, directed to you without reaching you, represents a proposal the other subject made you but you did not reciprocate.

The red lines represent your relations, and the black lines represent the relations between the other subjects.

On the right-hand side of the screen you can choose between two actions: **up** or **down** (you must choose one of them). Depending on your symbol and the decisions made by the subjects you linked to in the first stage, you can earn points. This is explained as follows:



If you are **circle** and you:

- choose **up**, you receive **4 points for each** of your connections choosing **up**
- choose **down**, you receive **2 points for each** of your connections choosing **down**

If you are **triangle** and you:

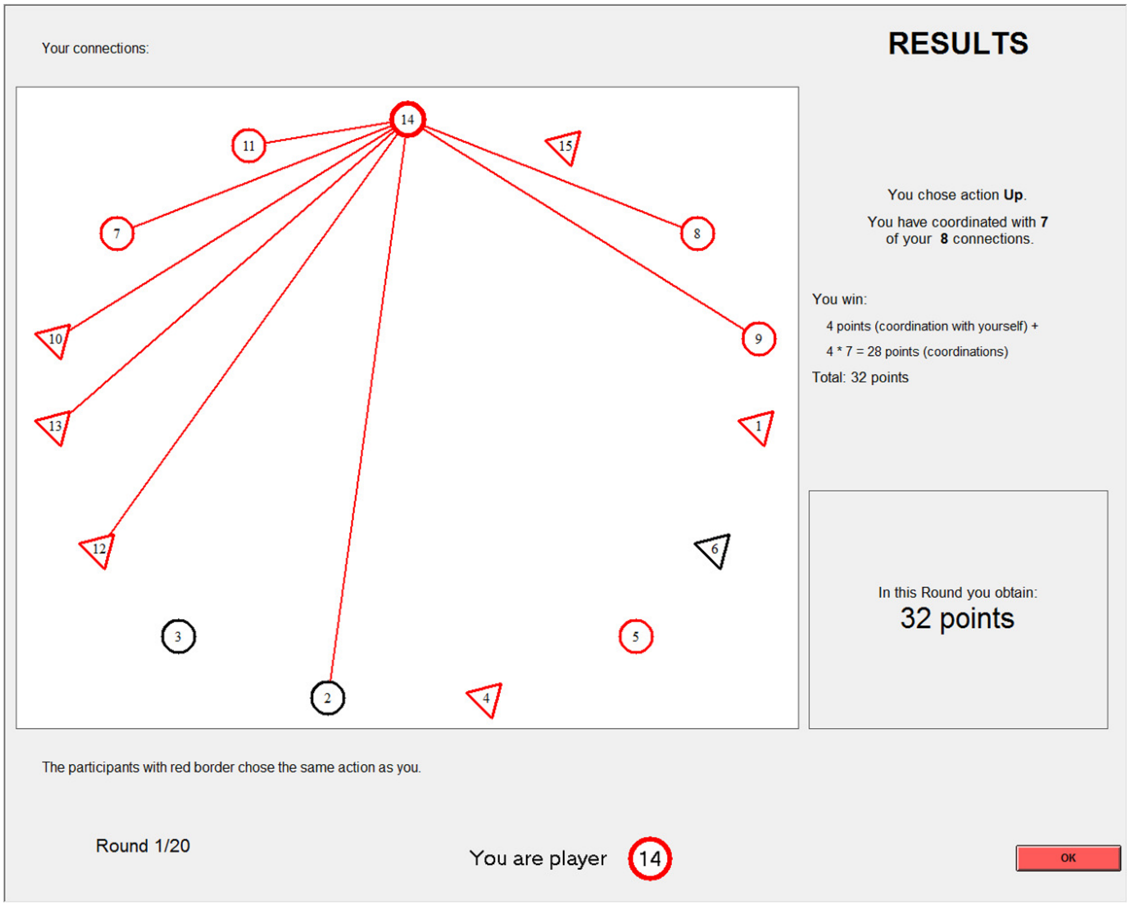
- choose **down**, you receive **4 points for each** of your connections choosing **down**
- choose **up**, you receive **2 points for each** of your connections choosing **up**

Phase 3. Earnings

In the last phase of each round you will see the points you earned given your interactions. On the left-hand side of the screen you will see the connections you formed. Those subjects choosing the same action as you will be displayed with a red border, otherwise they will have a black border. This will allow you to easily calculate the points you earn in the current round.

Please, bear in mind that you earn points for each subject you are linked to who chooses the same action as you (displayed with a red border). The exact amount of points (4 or 2) will depend on your symbol and the action you chose (as explained in Phase 2 (Action).

The total amount of points you earn will be the sum of the points you obtained during the 20 rounds of the experiment (the second stage).



Next, we present two examples:

Example 1: You are a circle, you are linked to 10 subjects, you have chosen *up* and 4 of your connections have chosen *up* as well (6 have chosen *down*). Therefore, you earn 4 points for coordinating with yourself (you always coordinate with yourself), and 16 (4×4=16) points for coordinating with the other 4. Your earnings in the round are 20 points in total.

Example 2: You are a circle, you are linked to 10 subjects, you have chosen *down* and 6 of your connections have chosen *down* as well (4 have chosen *up*). Therefore, you earn 2 points for coordinating with yourself (you always coordinate with yourself), and 12 ($2\times6=12$) points for coordinating with the other 6. Your earnings in the round are 14 points in total.

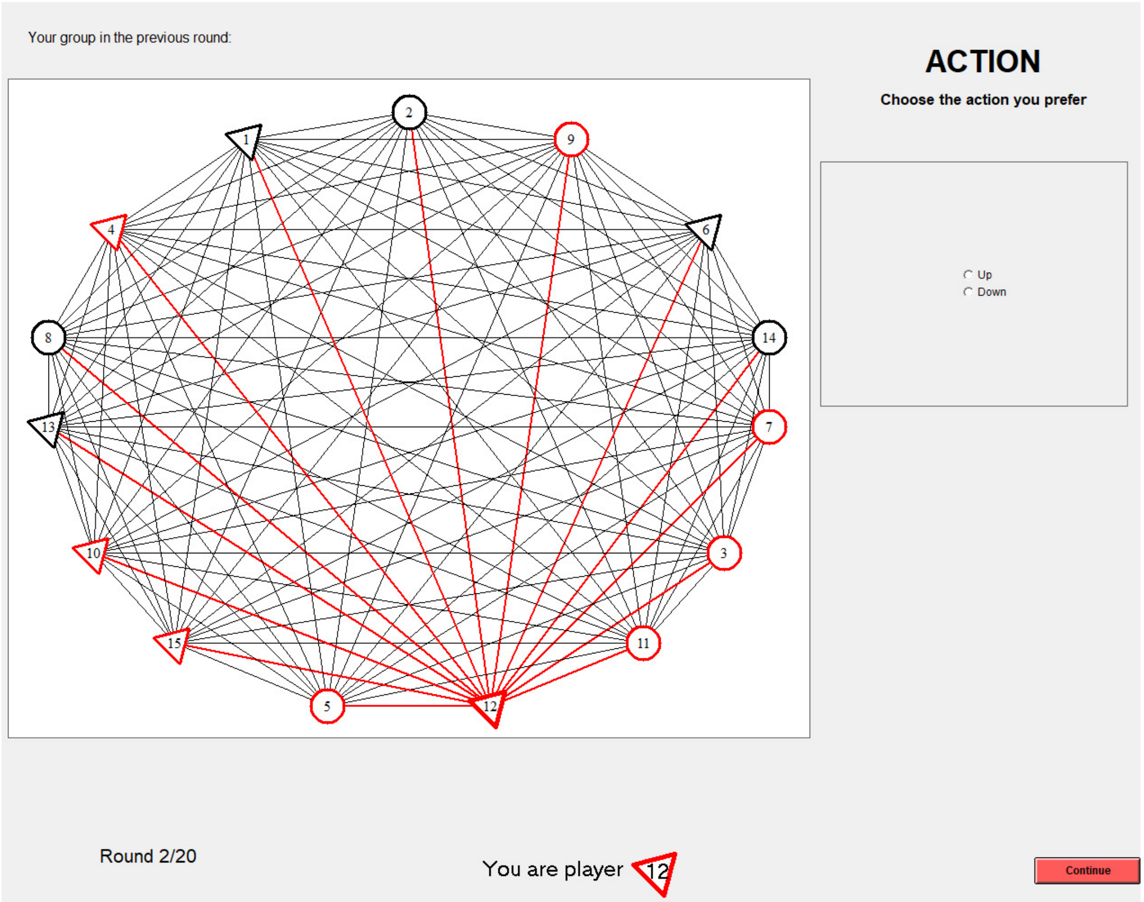
Specific to Treatment EXO only:

Each round consists of 2 phases: (1) Action and (2) Earnings.

Phase 1. Action

At the beginning of each round you will see the group of subjects you interact with and their choices in the previous round (in the first round you will see the subjects without any previous decision). You will see your number and your type, and the numbers and types of the other subjects, as illustrated in the image below. You will be highlighted with a thicker border, to facilitate that you can identify yourself in the screen.

On the right-hand side of the screen you can choose between two actions: **up** or **down** (you must choose one of them). Depending on your symbol and the decisions made by the subjects you linked to in the first stage, you can earn points. This is explained as follows:



If you are **circle** and you:

- choose **up**, you receive **4 points for each** of your connections choosing **up**
- choose **down**, you receive **2 points for each** of your connections choosing **down**

If you are **triangle** and you:

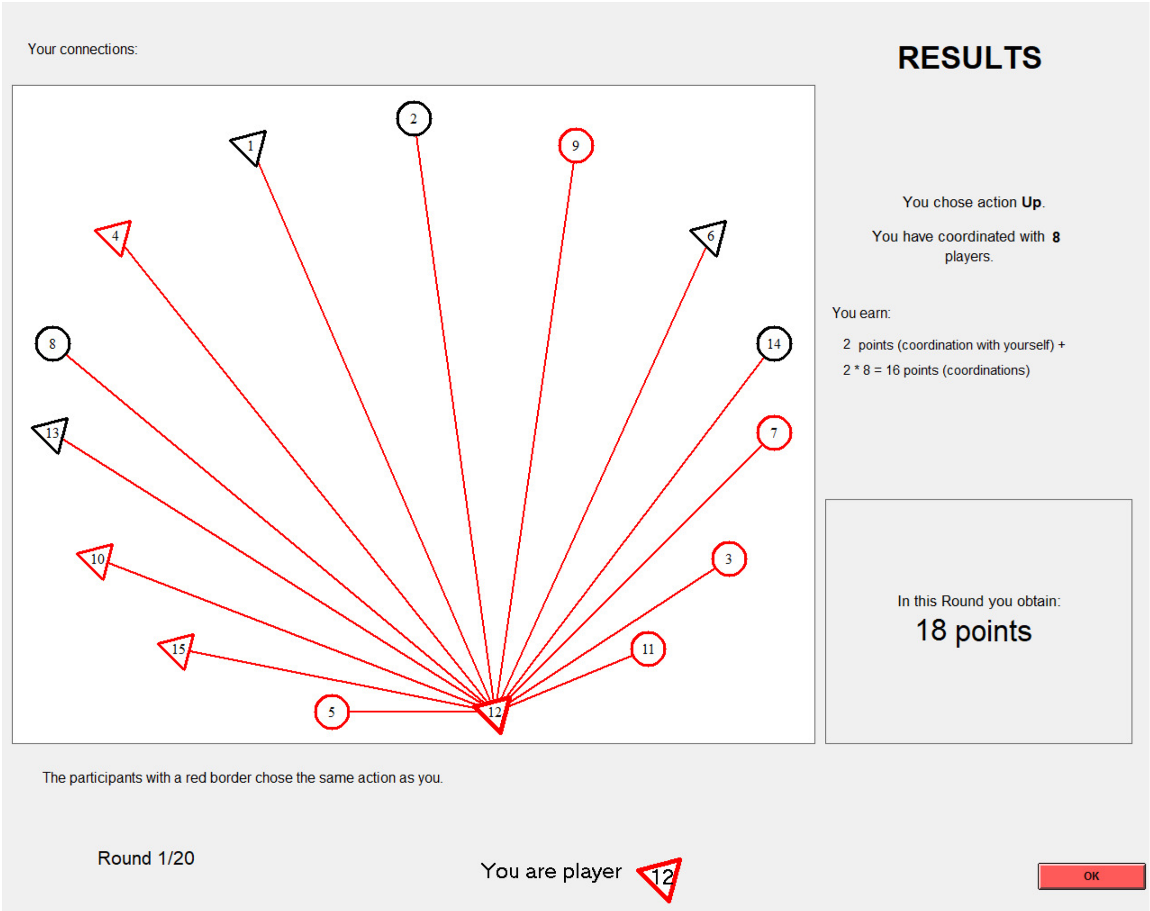
- choose **down**, you receive **4 points for each** of your connections choosing **down**
- choose **up**, you receive **2 points for each** of your connections choosing **up**

Phase 2. Earnings

In the last phase of each round you will see the points you earned given your interactions. On the left-hand side of the screen you will see the connections you formed. Those subjects choosing the same action as you will be displayed with a red border, otherwise they will have a black border. This will allow you to easily calculate the points you earn in the current round.

Please, bear in mind that you earn points for each subject you are linked to who chooses the same action as you (displayed with a red border). The exact amount of points (4 or 2) will depend on your symbol and the action you chose (as explained in Phase 1 (Action).

The total amount of points you earn will be the sum of the points you obtained during the 20 rounds of the experiment (the second stage).



Next, we present two examples:

Example 1: You are a circle, you have chosen *up* and 4 other subjects have chosen *up* as well (10 have chosen *down*). Therefore, you earn 4 points for coordinating with yourself (you always coordinate with yourself), and 16 (4×4=16) points for coordinating with the other 4. Your earnings in the round are 20 points in total.

Example 2: You are a circle, you have chosen *down* and 10 other subjects have chosen *down* as well (4 have chosen *up*). Therefore, you earn 2 points for coordinating with yourself (you always coordinate with yourself), and 20 (2×10=20) points for coordinating with the other 6. Your earnings in the round are 22 points in total.

All treatments:

Summary In each round, you can create connections. You will earn points from those subjects you are connected to who chose the same action as you (coordinate with you). The session consists of 2 stages, the first is a trial stage, which lasts 5 rounds, and the latter is the experiment and lasts 20 rounds. You will participate with the same 15 subjects for a whole stage (trial or experiment), but your group, symbol and number, and those of the other subjects, might change between stages.