Integration and Diversity

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

March 26, 2018

Abstract

Individuals prefer to coordinate with others but they differ on their preferred action. We consider both a setting in which interactions are exogenous and one in which individuals choose links that determine the interactions. Theory is permissive in both settings: conformism (on a single action) and diversity (with different groups choosing their preferred actions) are both sustainable in equilibrium. So we turn to laboratory experiments.

In an exogenous complete network, subjects choose to conform to the majority's preferred action. By contrast, when they choose interactions, individuals form dense networks but choose diverse actions. The freedom to link with others allows individuals to differentiate themselves and this greatly reinforces diversity. Standard theories – such as stochastic stability, team reasoning, k-level reasoning, inequity aversion – cannot account for this pattern of behavior. We suggest that players use linking to communicate the action they will play in the coordination game.

*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
‖Grupo Interdisciplinar de Sistemas Complejos, Departamento de Matemáticas, Institute UC3M-BS of Financial Big Data, Universidad Carlos III de Madrid, and Institute for Biocomputation and Physics of Complex Systems (BIFI), Universidad de Zaragoza, Email: anxo@math.uc3m.es.
This paper has been supported by the EU through FET-Proactive Project DOLFINS (contract no. 640772) and FET-Open Project IBSEN (contract no. 662725), and grant FIS2015-64349-P (MINECO/FEDER, UE). We are grateful to Marina Agranov, Gary Charness, Vince Crawford, Sihua Ding, Matt Elliott, Edoardo Gallo, Jonathan Newton, Theo Offerman, Gustavo Paez, 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]

Diversity in norms, values, and modes of behavior is valued – both for intrinsic and instrumental reasons – but it is also viewed as a social challenge. Academic work as well as popular writing has voiced a concern on the fragmenting of society along the lines of personal and social identity.[1] In the domains of language, food, dress, education and occupation, the returns to an action are intimately related to what others – especially those close to us – choose. Personal and social identity creates expectations on the preferred course of action in these domains. Thus in our day to day life we are confronted with a range of decision problems that share the following features: individually we would prefer to coordinate our action with the choices of others and at the same time we typically have differing notions of what is the best course of action. The goal of the paper is to better understand how we navigate this challenge.

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. There exist three equilibria: everyone conforming to a single action, up or down, and diversity with 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). Broadly, the outcomes take two forms: in one case, 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 case arises when 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

[1] For a classic early study of segregation in structured populations, see Schelling [1978]. For an overview of recent arguments on how identity affects politics in a liberal democracy, see Fukuyama [2006].

action maximizes aggregate welfare. Thus there is a multiplicity in outcomes in both the exogenous and the endogenous linking case and there is a tension between diversity and aggregate welfare.[2] We conduct laboratory experiments to better understand how players choose actions and how these choices are affected by whether the network is exogenous or if it is 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). In all, there are 6 groups with exogenous, and 6 groups with endogenous linking. We find that, with exogenous interaction, conformity on the majority 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 choose diversity. Thus the freedom to create links has a powerful effect on behavior and on aggregate welfare.

We show that standard theories – such as stochastic stability, team reasoning, k-level reasoning, and inequity aversion – cannot account for this evidence. We take the view that in these contexts, individuals look for cues in the environment and mechanisms to help them resolve very complex coordination problem. Following up on this broad idea we explore the idea that in the endogenous linking setting players use their linking choices as a way to convey information about which action – up or down – they intend to choose.

We first examine the role of linking costs. We consider both positive and negative costs, but the idea with negative linking costs (or link subsidy) is easier to appreciate and we discuss it in detail here. As cost of a link is negative, for a minority player to *not* form a link with a majority player is a costly action. So we interpret not forming a link as conveying the message that a player is minded to not conform with the majority. In particular, in the game with negative linking costs, we show that there exist two equilibrium outcomes: *integration with conformity* and *integration with diversity*. In the experiment, we find that in all 6 groups, subjects form almost fully connected networks (on average 100 out of 105 possible links) and they choose diversity. An examination of the data reveals that while individuals form all links within their own type, some across type links are missing. Moreover, for a minority player, the ratio of across links to within type links is positively correlated with him choosing the majority action, and we find evidence that such a ratio significantly affects other minority players' choice to conform.

[2]In Section 5, we discuss a number of alternative equilibrium selection models.

Secondly, we study behavior of subjects in exogenous networks that correspond to the networks created in the treatment with endogenous linking and zero costs. The thought here is to look directly at the signalling role of links: if the act of choosing links *per se* does not play a communication role, then behavior in exogenous networks should be the same as behavior in the endogenous treatment. We consider two network configurations reflecting different levels of symmetry in missing connections across preference types. We conduct experiments with 6 groups for each network configuration, so there were 12 groups in all. We observe that diversity is obtained in only 7 out of 12 groups, i.e., in 58% of the cases. By contrast, recall that in the endogenous treatment subjects chose diversity in all 6 groups, i.e., in 100% of the cases. In our view, this supports the view that the pattern of linking acts as a signal of intention on action choice in the coordination game.

Our paper is a contribution to the study of social coordination. Following the early contributions of Schelling [1960] and Lewis [1969], there has been a large and influential strand of research on coordination problems in economics. 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.

There is a large experimental literature on social coordination, see e.g., Charness et al. [2014], Crawford [1995], Isoni et al. [2014], Kearns et al. [2012]. Our experimental work departs from this work in that it brings together heterogenous preferences on actions and we allow for individuals to choose with whom to interact. Bringing together these two features has large effects. To bring this out clearly, consider the minimum effort game: it offers a simple way for thinking about situations in which everyone must agree about the outcome and yet there is a range of Pareto ranked (equilibrium) actions. The early experiments on this game showed that subjects converged to the lowest welfare Nash equilibrium [Van Huyck et al., 1990]. A number of variations on the original experiment with varying outcomes have been reported since then; notable contributions include van Huyk et al. [1991], Crawford and Broseta [1998]. Our paper is related to a recent paper by Riedl et al. [2016] who introduce the possibility that players can choose their partners while playing the minimum effort game. They find that endogenizing the choice of partners has a dramatic effect on behavior: players now 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.

At a more general level, our paper also contributes to the work on identity. There is a large literature on identity, spanning across several disciplines in the social sciences and in philosophy. In recent years, there has been a great deal of interest in understanding the ways in which identity shapes behavior in society, organizations, markets, and in local government, see e.g., Advani and Reich [2015], Akerlof and Kranton [2000], Alesina et al. [1999], Bisin and Verdier [2000], and Sethi and Somanathan [2004]. Types in our setting may naturally be interpreted as an aspect of identity. In particular, following the work of Sherif et al. [1988], a number of papers have looked at the role of identity in shaping behavior in an experimental setting. The papers have developed an experimental design in which identity is 'minimal': individuals are made to associate themselves with others who share a similar view on something orthogonal to the experiment itself. A common example is shared ideas on a piece of art: so two individuals share the same identity if they like the same painting and not otherwise. The experiment then shows how this 'minimal group' identity can play a large role in shaping behavior in games and decision problems. A leading paper in this line of work, Chen and Chen [2011] shows that group identity has direct effects on social preferences, which in turn can induce higher effort in the minimum effort game. They show that exogenously varying the salience of identity leads to a significant improvement in efficiency of play in this game.

Relative to Chen and Chen [2011], an important difference is that we allow for heterogenous preferences. In our model, 'identities' are reflected in payoff differences and they are kept constant across the exogenous and endogenous linking treatments. Our principal experimental finding is that endogenous linking allows distinct preferences more space to become salient. This is perhaps best revealed in the treatment where the costs of linking are zero. Now linking with everyone is (in a loose sense) a 'weakly dominant' strategy, and so subjects should create a complete network. But then we are in the same setting as the exogenous networks, and so subjects should all conform on the majority's preferred action. In the experiment, however, subjects create 'almost' complete networks but different types nevertheless choose their own preferred actions! Thus the freedom to choose links helps individuals differentiate along preference types.

The paper is organized as follows. Section 2 presents the model and the theoretical analysis. Section 3 presents our experimental design and the experimental findings on endogenous versus exogenous networks. Section 4 develops our argument on the role of links as signals. Section 5 explores four alternative theoretical approaches — stochastic stability, team reasoning, social preferences, and k-level reasoning – to explain our findings. Section 6 concludes. Appendix A contains some of the proofs, Appendix B contains some additional experiments while Appendix C 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 total function x_i mapping every undirected network $\overline{g}$ that can result from the first stage to an action in $A = \{up, down\}$. Formally, $x_i : \overline{G} \rightarrow A$, 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 less preferred action) with $\beta < \alpha$. This payoff function is taken from Ellwardt et al. [2016].

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. The following result, taken from Ellwardt et al. [2016], characterises equilibrium behavior in an arbitrary network.

Proposition 1. *Fix a network g. A strategy profile x^* is a Nash equilibrium if and only if, for every $i \in N$:*

$$x_i^* \begin{cases} = \theta_i & \text{if } |\chi_i(\overline{g})| > \frac{\beta}{\alpha+\beta}|N_i(\overline{g})| - \frac{\alpha-\beta}{\alpha+\beta} \\ \neq \theta_i & \text{if } |\chi_i(\overline{g})| < \frac{\beta}{\alpha+\beta}|N_i(\overline{g})| - \frac{\alpha-\beta}{\alpha+\beta} \end{cases}$$

The proof of this result follows from computations which are presented in the main text as they provide a good sense of the basic trade-offs involved. 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 result 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 2. *Fix a complete network g. Suppose x^* is a Nash equilibrium. Then either (i) Every player selects the same action. This is possible if $n \geq \alpha/\beta$, OR, (ii) every player selects their preferred action. This is possible if $|N_u|, |N_d| \geq \frac{\beta(n+1)}{\alpha+\beta}$.*

We sketch the proof here. To fix ideas, consider conformism on the majority action. 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 conformism is an equilibrium if $n \geq \alpha/\beta$. Similar computations also hold for the conformism on the minority action equilibrium.

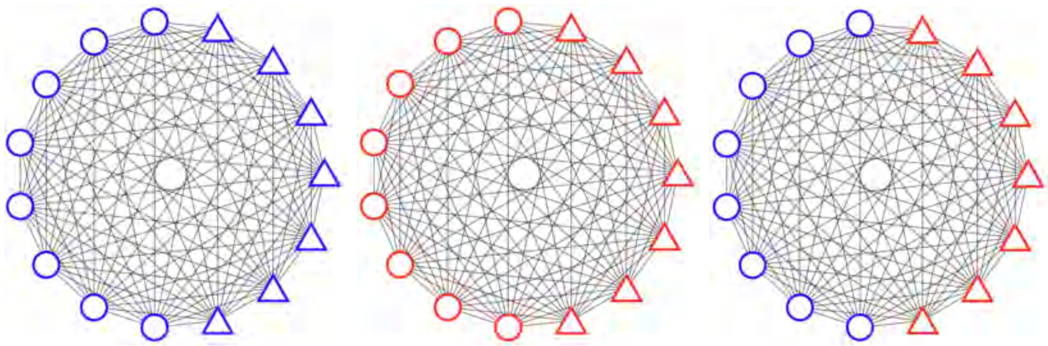
Turning to the diversity outcome, note that if some player i benefits by playing $x_i \neq \theta_i$, then so would every player j of the same type. It then follows from Proposition 1 that diversity is an equilibrium if:

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

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

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

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



(a) Conformity (majority) (b) Conformity (minority) (c) Diversity

Figure 1: Nash equilibria in complete network

So 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. There are 8 players represented by *circles* and the remaining 7 individuals are represented by *"triangles"*. The circles prefer action 'blue', while the triangles prefer 'red'.

We now solve for the two stage game with link formation and action choices. We adapt the pairwise stability notion from Jackson and Wolinsky [1996] to our setting. So, 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 is such that $x_{-ij}(\overline{g} - \overline{g}_{ij}) = x_{-ij}(\overline{g})$, and $x_l \in \arg\max_{x'_l \in X_l} u_l(\theta_l, x'_l, x_{-l}, \overline{g} - \overline{g}_{ij})$ for $l \in \{i, j\}$.*
- *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 is such that $x_{-ij}(\overline{g} + \overline{g}_{ij}) = x_{-ij}(\overline{g})$, and $x_l \in \arg\max_{x'_l \in X_l} u_l(\theta_l, x'_l, x_{-l}, \overline{g} + \overline{g}_{ij})$ for $l \in \{i, j\}$.*

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 only require that the players directly affected by a change of the link re-optimize actions; all other players remain with their pre-specified equilibrium action, corresponding to network $\overline{g}$. This restriction to very local action adjustments are in the spirit of pairwise stability, but they do not fully reflect the idea behind sub-game perfection. Our aim here is to show that conformism 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.

A useful implication of Definition 1 is that in a pairwise stable network-action pair $(\overline{g}, x(\overline{g}))$, if $k = 0$, then any two players who choose the same action in the second stage must also be linked with each other: *for any pair $i, j \in N$, $x_i(\overline{g}) = x_j(\overline{g})$ only if $\overline{g}_{ij} = 1$.*

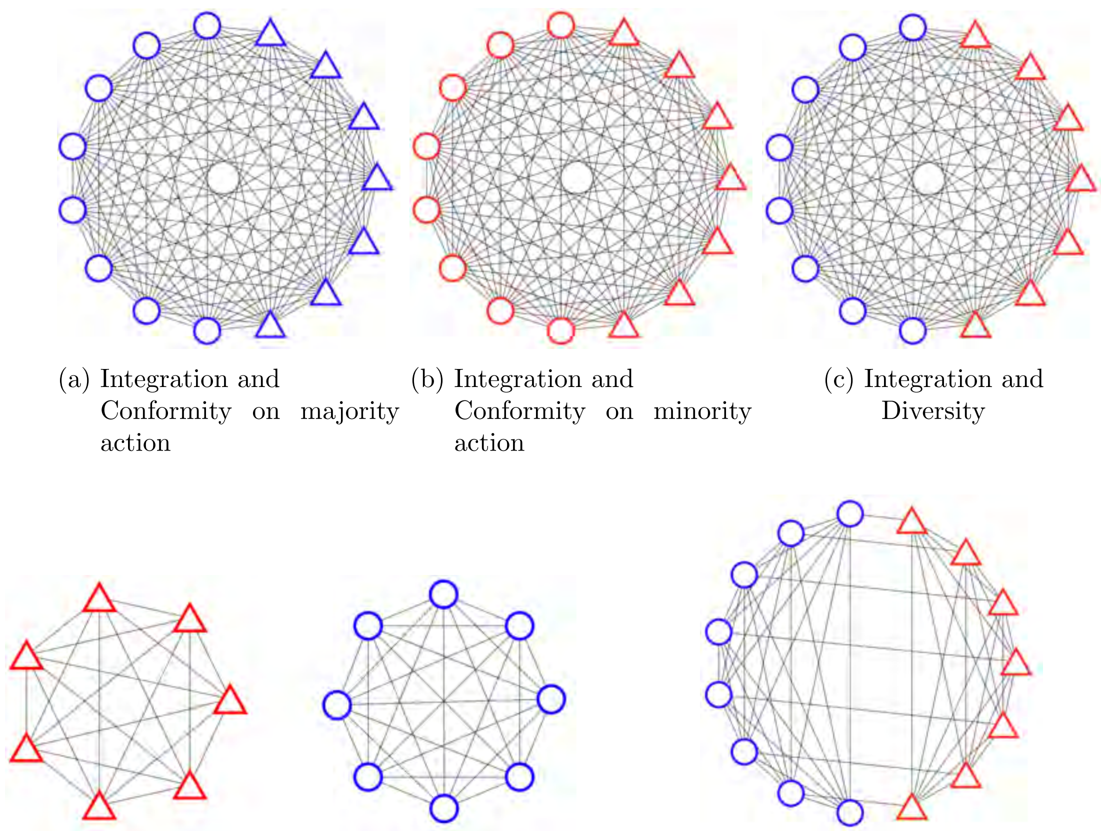
Proposition 3. *Suppose $k = 0$ and $(\overline{g}^*, x^*(\overline{g}^*))$ is pairwise stable. Then:*

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

(ii) $\overline{g}^*$ is partially connected. This allows for two distinct completely connected components as well as a connected but incomplete network. In the former case, individuals within a component choose the same action but actions differ across components. In the latter case, individuals who choose the same action are connected while some of the links across different action individuals are missing.

The proof of Proposition 3 is immediate from the observation preceding it. This result highlights four types of equilibrium outcomes. *integration with conformity* arises when the network is complete and everyone chooses the same action. *Integration with diversity* arises when the network is complete and everyone chooses their preferred action. *Segregation with diversity* arises when the network contains two components and individuals in the two components choose a different action. Finally, *partial integration with diversity* arises when individuals choose distinct actions, all individuals with the same action are linked while the agents choosing dissimilar actions are only partially linked.

We illustrate these outcome 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 and partial integration are illustrated in the bottom half of Figure 2.



(a) Integration and Conformity on majority action

(b) Integration and Conformity on minority action

(c) Integration and Diversity

(d) Segregation and Diversity

(e) Partial integration and Diversity

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

We now turn to social welfare: we define aggregate welfare as the sum of earnings 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 conformism on the majority action maximizes social surplus.

Proposition 4. *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 integration and conformity on the majority's preferred action.*

The proof is presented in Appendix A. The result says that diversity is never socially desirable. To develop some intuition for the result, consider the complete network. Fixing the behavior of one group, it is never desirable for the other group to mix actions. This follows from the coordination externalities inherent to our model. So we only need to compare the two outcomes: one, where everyone conforms to action up and the other where everyone conforms to action down. The concluding step then shows that conformism on up is better if and only if the group that prefers up constitutes a majority. So, in our example, with exogenous complete network, the socially efficient outcome corresponds to Figure 1(a). Similarly, in the endogenous links treatment, the unique socially efficient outcome is presented in Figure 2(a).

We summarize the theoretical analysis as follows: in the exogenous complete network there exist multiple equilibria exhibiting conformity and diversity. The conformity equilibria are independent of group sizes, while the diversity equilibria can only arise if the minority group is not too small. With endogenous links, there exist multiple equilibria exhibiting full integration with conformity, segregation with diversity, and also partial and full integration with diversity. 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.

3 Experiments

3.1 Experimental design

To evaluate the effects of linking on coordination and on welfare, we study two main treatments: **ENDO** and **EXO**. The treatment **ENDO** starts with an empty network and

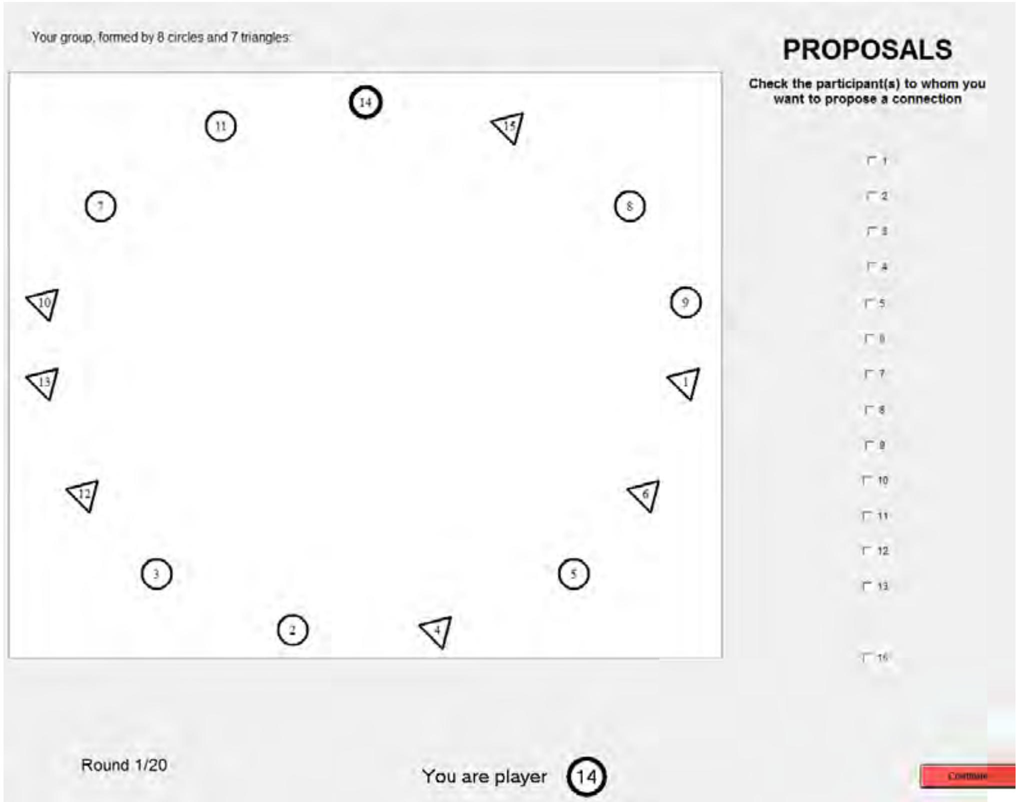


Figure 3: Individual Types

refers to the two stage model of linking and action choice. The treatment **EXO** specifies that players are located in an exogenously given complete network and they simply choose between two coordination actions.

Throughout we consider groups of 15 subjects. Subjects interact repeatedly, within the same group, for 20 rounds (plus 5 unpaid trial rounds). Prior to the start of play, subjects are informed of a symbol, either a circle or a triangle, and an identification number, from 1 to 15, assigned to them. Every subject 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 session. Groups are composed of 8 circles (the majority group) and 7 triangles (the minority group). Figure 3 presents the screen that subjects see at the start of the experiment (note that the positions of circles and triangles are mixed to avoid potential visual biases).

In the treatment **ENDO**, there are two stages. First subjects simultaneously make

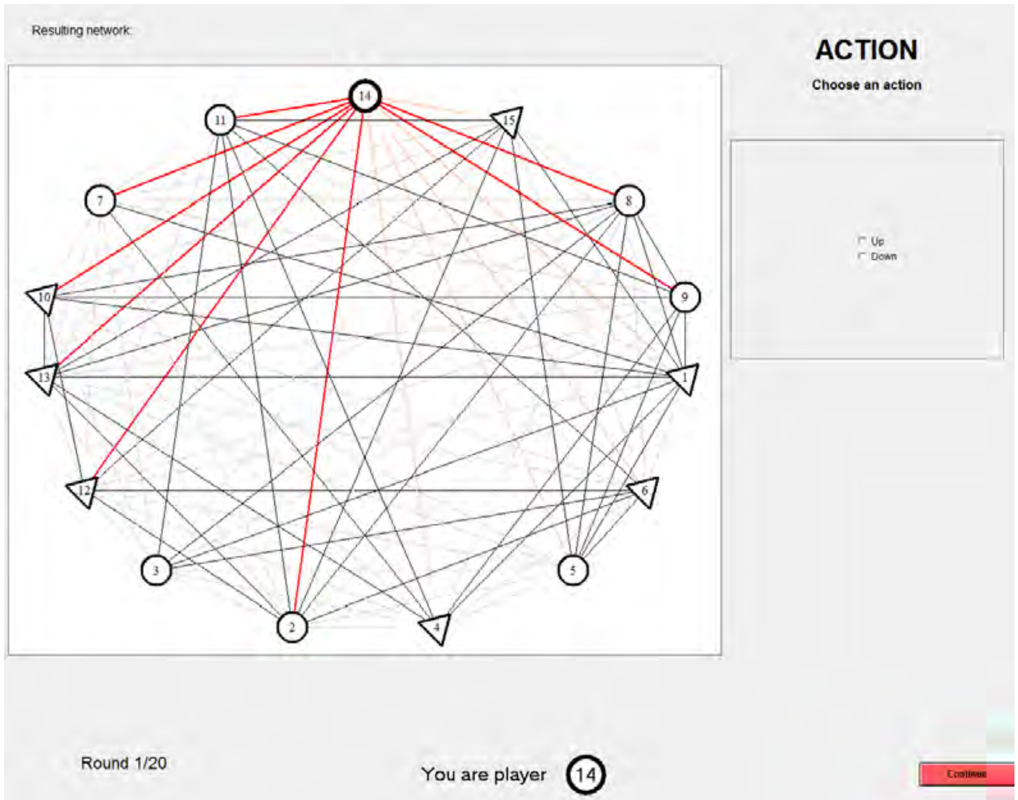


Figure 4: Linking proposals: reciprocated and unreciprocated

proposals to a subset of the others in their group. Reciprocated proposals lead to the creation of links. No cost is paid for any link formed (i.e., $k = 0$). Then, in the second stage, subjects are informed of the links proposed and those that are formed in stage 1. After observing the created network, subjects choose one of two actions: *up* or *down*. Figure 4 illustrates information about the network that they observe at this point. In the screenshot, links that are proposed but not reciprocated are represented as light 'incomplete' edges,[3] while reciprocated (bilateral) proposals are represented as dark 'complete' links. So in the screenshot, player 14 creates 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 values of the parameters are $\alpha = 4$ (payoff for coordinating with a connected player on one's *most* preferred action), $\beta = 2$ (payoff for coordinating with a connected player

[3] An edge departing from node i towards node j without connecting j means that player i forms a link with player j while j does not form a link with i.

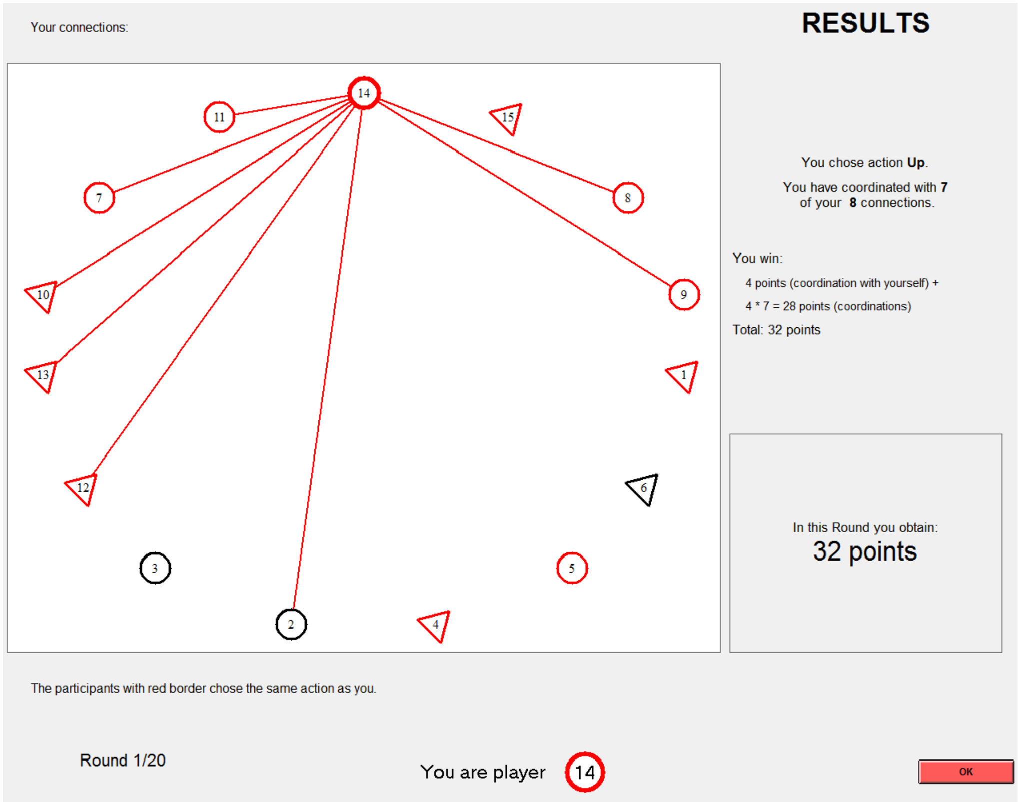


Figure 5: Feedback from previous round

on one's *least* preferred action), and $k = 0$ (cost of any bilateral link). For a subject with symbol circle (triangle), his preferred action is up (down). Every player sees the outcome of the game on the screen and his net payoffs as in Figure 5. Here we see 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 round r>1, in stage 1, every player receives information about every other player's links and actions and his own payoff, as shown through Figure 6.

In the treatment **EXO**, all subjects interact with every other group member in a complete network. The subjects are shown the complete network and they have to choose between actions *up* and *down*.[4] Given that there is no linking decision, there are also no

[4]The complete network is however shown as it would be in **ENDO**, had the complete network emerged. See the instructions in Appendix C.

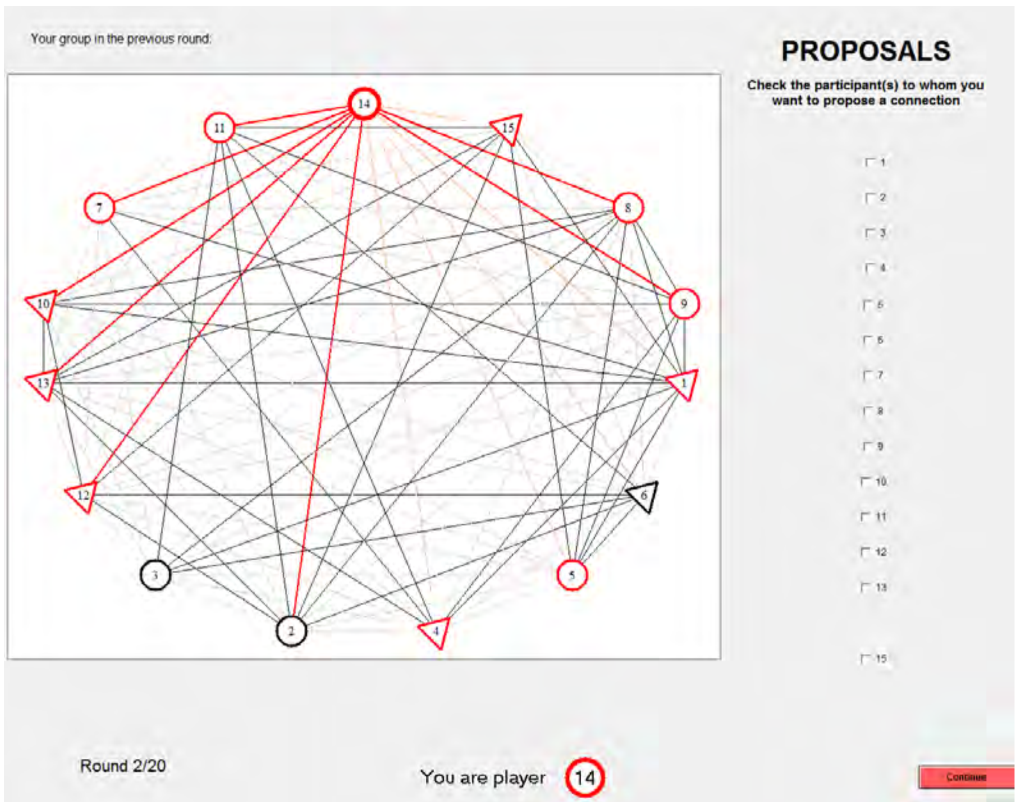


Figure 6: Neighborhood, Actions and Payoffs

linking costs. For this reason, and to make earnings comparable between treatments, the parameters in **EXO** are $\alpha = 4$ and $\beta = 2$. The instructions handed out to subjects are presented in Appendix C.

3.2 Experimental procedure

The experiment was conducted in the Laboratory for Research in Experimental and Behavioural Economics (LINEEX) at the University of Valencia. Subjects interacted through computer terminals and the experiment was programmed using z-Tree [Fischbacher, 2007]. 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. At the end of the experiment each subject answered a debriefing questionnaire. The standard conditions of anonymity and non-deception were implemented in the experiment.

Subjects were recruited through an online recruitment system. For each treatment, we conducted 2 sessions; in each session there were three groups with 15 subjects each. Thus there were six groups per treatment. Each session lasted between 90 and 120 minutes, and on average subjects earned approximately 18 euros.

3.3 Experimental Findings

For ease of exposition, we will present average behavior across groups on a round by round basis in the various plots. However, as the groups are playing a repeated game across twenty rounds, clearly observations across rounds for a group are not independent. So we will simply take the average across the twenty rounds for each group as the observation. This means that in the statistical tests we will have six independent observations (corresponding to the six groups), per treatment.

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 (that are not equilibrium in the static one shot game). 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, we feel that repeated game effects are not

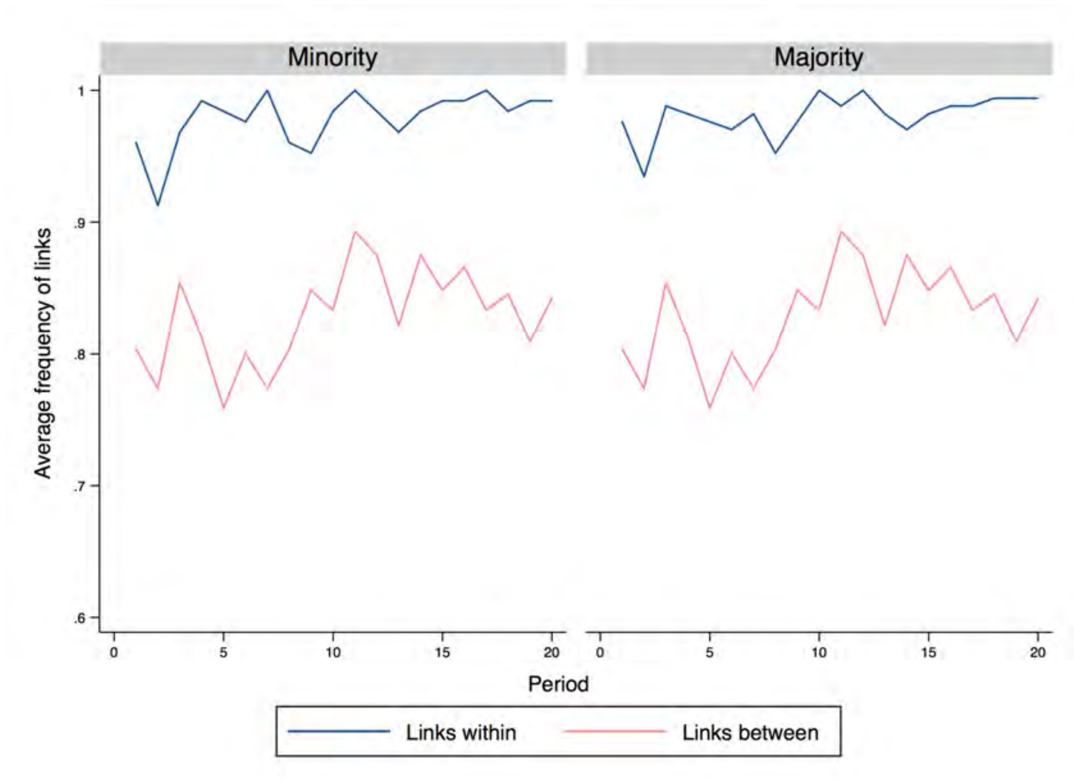


Figure 7: Link Choices in **ENDO**.

central to understanding this difference.

Consider the treatment **ENDO**: Figure 7 shows that the network is highly connected from round 1, and that the high rates of connectivity continue over time, without much variation. Subjects create roughly 94.5 links out of a maximum total of 105 links; in other words, individual degree are on average 12.59 (out of a possible 14). So the difference in network density between **EXO** and **ENDO** is very small. A first thought is that since the networks under the two treatments are so similar, the payoff trade-offs facing subjects too are similar and so we should expect that subjects choose the same actions in the coordination game under both treatments.

However, our data show that under **EXO**, in five out of the six groups, all subjects conform fully with the majority action, while under treatment **ENDO**, individuals in all six groups choose actions in line with their preferences and diversity obtains.[5] In particular, the average number of subjects choosing the majority's action across rounds is significantly lower under **ENDO** as compared to **EXO**, $8.18 < 12.68$ (Wilcoxon-Mann-Whitney: $z = 7.73$, $p < 0.0001$). Moreover, using the Jonckheere-Terpstra test we observe there is a significant difference in the speed of convergence to full conformity for the minority players in the two treatments:[6] Under **EXO**, the increasing trend toward conformity is only present in the first ten rounds (T_{JT}=2.26, p=0.02), but there is no variance (and thus no trend) in the latter ten rounds. Under **ENDO**, the level of conformity has no trend in the first-ten periods (T_{JT}=29.5, p=0.19) and significantly falls away from conformity in the last ten periods (T_{JT}=35, p=0.015). Figure 8 shows that the main source of the difference between the treatments is the choices of the minority.

To summarize: under both treatments, the majority chooses its preferred action almost from the start and persists with it across all the the rounds. The behavior of the minority is dramatically different depending on whether links are exogenous or endogenous. Under **EXO**, around 40% of the minority start by conforming to the majority action and by round 10 this fraction is well in excess of 80%. Under **ENDO**, the most of minority individuals – around 90% – choose their preferred action from the start, and by round 10 this goes up to 95% of the group.

Experimental Finding 1. *In the exogenous complete network setting, subjects conform to the majority's preferred action. By contrast, in the endogenous linking game, subjects create an (almost) complete network and choose diversity (along their preferred actions).*

This suggests that allowing individuals the freedom to choose links with others leads to dramatically different behavior in the coordination game. We take the view that individuals are facing a very complex coordination problem, due to the combination of many players and the heterogeneity in preferences. So it is only natural that they will try and use cues from the environment and instruments that they have available to simplify the coordination problem. The experiment points to the role of linking.

[5] We note that in the unique non conforming group from **EXO**, the minority and the majority choose their preferred action. There are no differences in results regarding the tests used, except for the last case: the number of minority players conforming in **EXO** is significantly different from 7 in all rounds if the outlier group is included. Therefore, we omit this group from the analysis and in the illustration in Figure 8.

[6] This is a non-parametric test for ordered differences of a response variable among classes. We are testing here the null hypothesis that the distribution of frequency of the level of conformity does not differ across periods. The alternative hypothesis is that there is an ordered difference among rounds.

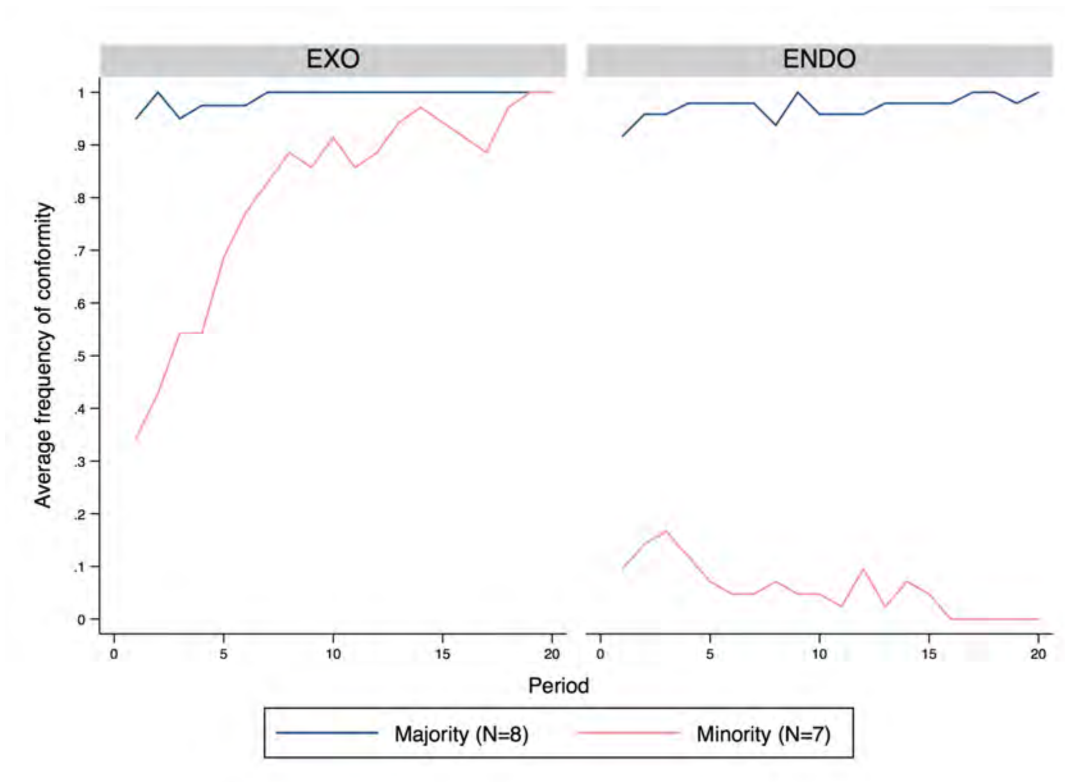


Figure 8: Action choice in **ENDO** and **EXO**.

Standard theories – such as stochastic stability, team reasoning, k-level reasoning, inequity aversion – cannot account for the sharply different behavior in the exogenous and endogenous network settings. Section 5 presents a detailed analysis of the implications of these theories and why they are inadequate. We now explore the broad idea that players use their links as a way to signal which action (up or down) they intend to choose.

4 Links as a Communication Device

First we will look at the role of linking costs. Take the case of negative costs: for a minority individual to not form a link with a majority preference type individual is a costly action and so we interpret that as conveying the message that they are minded to not conform with the majority. As not forming a link is costly, individuals should choose to not form only a minimum number of links (removing more links increases the strength of the signal but decreases immediate payoffs). In the positive cost case, 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. So not forming a link signals an intention to stick to one's own preferred action.

Second, we will sample networks from the endogenous treatment, fix them as exogenous networks and examine behavior. The interest here is in seeing whether the behavior of subjects remains unchanged or if it is different from the behavior in the endogenous network. If behavior is markedly different then that would suggest that the act of linking *per se* is important.

The following table provides a summary of the experimental design in this section.

Network ($N = 15$)			
Endogenous		**Exogenous** (incomplete networks)	
$k = -0.3$	$k = 0.5$	-	
SUBS (6 groups)	**COST** (6 groups)	**EXOSYM** (6 groups)	**EXOASYM** (6 groups)

Table 1: Experimental Treatments

The negative linking cost treatment is denoted as **SUBS**, while the positive cost treatment is denoted as **COST**. The two exogenous network treatments take up different patterns of missing links: when the missing links are evenly distributed across the minority individuals we get the treatment **EXOSYM** and when they are unevenly distributed we get the treatment **EXOASYM**.

4.1 Costs of Linking

We start with the case where linking has a negative cost. Observe that any two players can strictly increase their payoffs by forming a link, regardless of whether they subsequently coordinate their actions. So it follows from Definition 1 that in a pairwise stable network-action pair, *for any pair* $i, j \in N$, $\overline{g}_{ij} = 1$. Then the following result is immediate:

Proposition 5. *Suppose* $k < 0$ *and* $(\overline{g}^*, x^*(\overline{g}^*))$ *is pairwise stable. Then:*

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

(ii) $\overline{g}^*$ *is complete and diversity obtains,* $\forall i \in N$, $x_i^*(\overline{g}^*) = \theta_i$.

Thus there exist only two types of pairwise stable outcomes: *integration with conformity* and *integration with diversity.*

In our experiment, we set the parameters as $\alpha = 4$, $\beta = 2$, and $k = -0.3$.[7]

Turning to the experiment, we present average behavior in Figure 9. The first observation is that connectivity is high and that it is higher than under the treatment **ENDO**, $101.4 > 94.5$ (Wilcoxon-Mann Whitney: $z = 7.631$, $p < 0.0001$). Both individuals create all the links with others of the same preference type and the few missing links all involve members of different preference type.

The left part of Figure 10 presents patterns of choice in the coordination game. We observe quick convergence to diversity in actions.

To summarize:

Experimental Finding 2. *When the cost of linking is negative, subjects choose an (almost complete) network and diversity of actions.*

We turn next to costly links. When links are costly, two players should only form a link if they intend to choose the same action in the coordination game. And, it follows from Definition 1 that *for any pair* $i, j \in N$, $x_i(\overline{g}) = x_j(\overline{g})$ *if and only if* $\overline{g}_{ij} = 1$. Building on this observation, we get the following result.

[7]Note that we do not have $\alpha - k = 4$ and $\beta - k = 2$ here. It can be shown that if $k < -1/4$, then setting $\alpha - k = 4$ and $\beta - k = 2$ means that integration with conformity is no longer Pareto dominant as in **EXO** and **ENDO**.

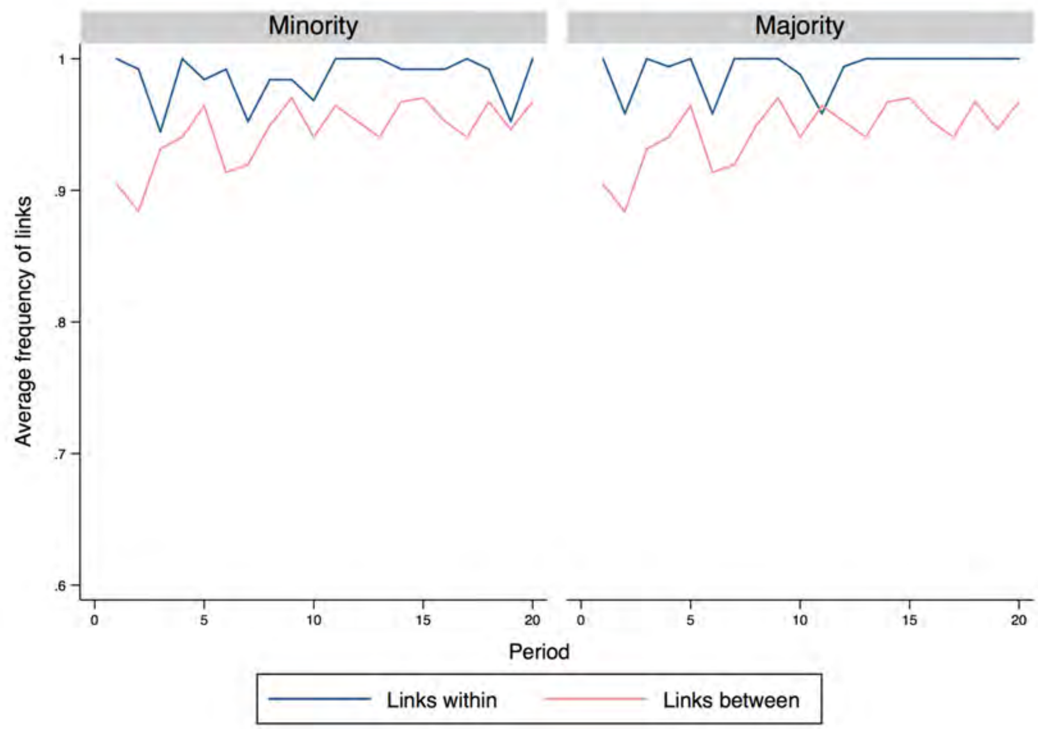


Figure 9: Link Choices in **SUBS**.

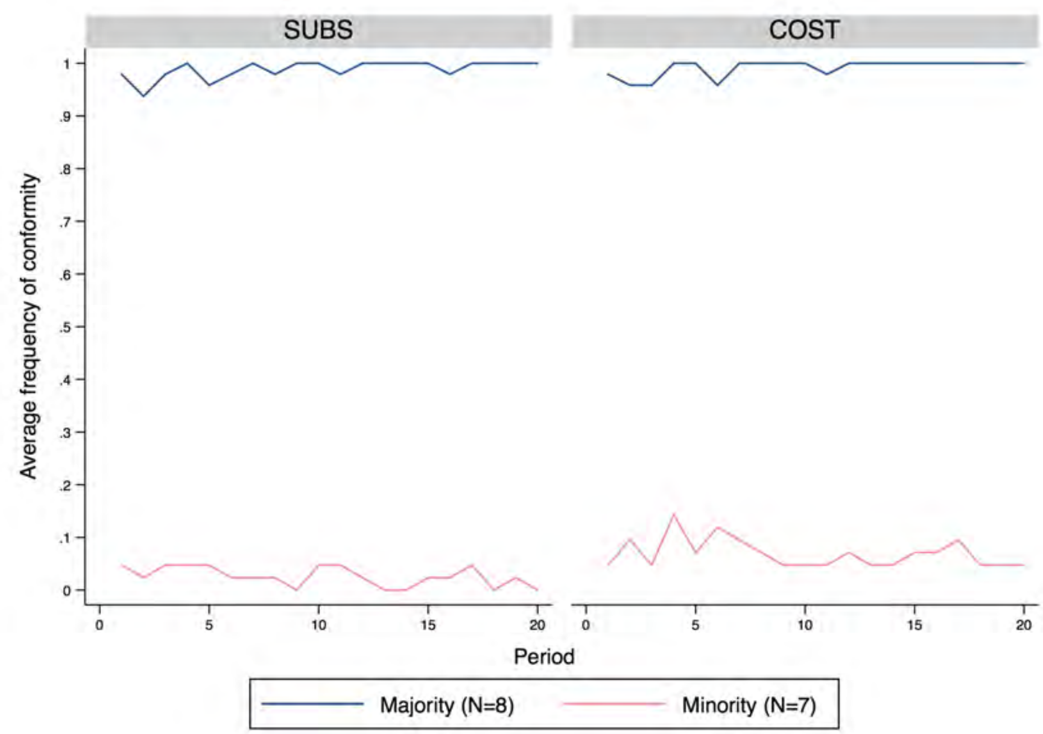


Figure 10: Action Choice in **SUBS** and **COST**.

Proposition 6. *Suppose $k > 0$ and $(\overline{g}^*, x^*(\overline{g}^*))$ is pairwise stable. Then:*

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

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

Thus, there are two types of pairwise stable outcomes: *integration with conformity* and *segregation with diversity*. In treatment **ENDO** (with $k = 0$), we observe that subjects created dense networks but choose diverse actions. As we raise the costs of linking our conjecture is that the pressure toward diversity will be reinforced.

In the costly links treatment **COST**, we set the parameters as $\alpha = 4.5$, $\beta = 2.5$ and $k = 0.5$.[8] We observe that individuals are active in creating links from early on, but this linking activity is mostly focused within types. Linking across players of different types is very limited at the start, and becomes rarer over time. Individuals choose their own preferred action: there is convergence to diversity.

At the start, more than 80% of the links within the groups are formed, but about 20% of the cross group links are also formed. By round 10, this tendency accentuates and around 90% of the within group links are in place, but less than 10% of the cross group links are being formed. Eventually, the network converges to two distinct complete components that have virtually no links between them. Out of the 105 links that can be formed, there are on average (across rounds) only 53. The average degree of the minority is 6.35 and that of the majority is 7.69.[9] We note that the majority form on average the same number of within-type links in **COST** and **ENDO** ($z = 1.596$, $p = 0.1105$), while the minority was significantly less connected in **COST** ($z = 5.292$, $p < 0.0001$). But the main difference is in the across-group ties: this decreases from 6.19 links in **ENDO** to 0.98 links in **COST** ($z = 7.699$, $p < 0.0001$). Thus we see the emergence of (almost) complete segregation, Figure 11.

We now turn to actions: Figure 10 illustrates the dynamics of action choice under **COST**. The main observation is that subjects choose diversity. To summarize:

[8] 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 those treatments.

[9] Recall there are 7 (8) players in the minority (majority) so that each of them can link to at most 6 (7) others sharing the same type.

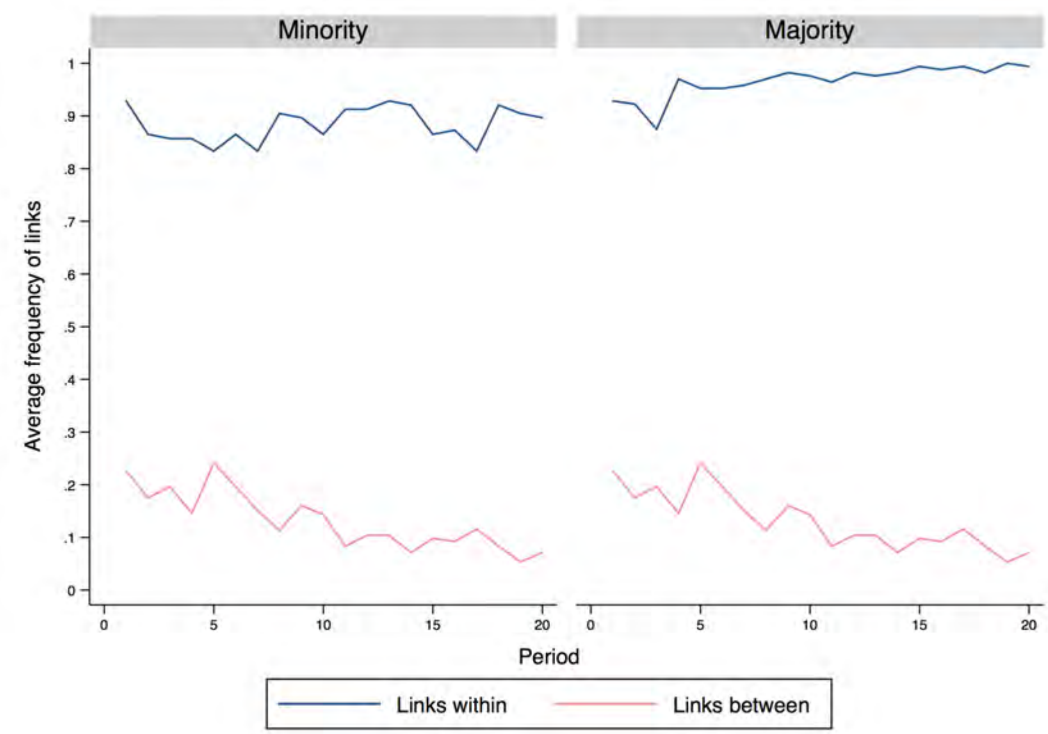


Figure 11: Link Choices in **COST**.

Experimental Finding 3. *When the cost of linking is positive, subjects create a segregated network and choose diversity of actions.*

How can we understand the linking patterns and the choice of actions in the positive and negative cost treatments? In the negative cost setting, it would be natural to hypothesize that a player would abstain from linking only if there were strategic advantages to doing so: they want to signal an intention of choosing their preferred action. At a more general level, this suggests that there should be a positive correlation between cross group linking and conformity with majority action. We run a regression of the probability that a player in the minority chooses to conform given the ratio of across versus within group links.

Formally, for every minority player i ($\theta_i = down$), we define:

$$Rate(i) = \frac{\sum_{j \in N: \theta_j \neq \theta_i} \bar{g}_{ij}}{\sum_{j \in N: \theta_j = \theta_i} \bar{g}_{ij}} \tag{7}$$

Table 2 presents multiple models pooling the data of all the treatments with endogenous linking. The key finding is that there is a positive and statistically significant correlation between the ratio of links with majority and the likelihood of conforming to the majority action. This effect is robust across treatments. Moreover, the ratio of links of others of the same type also significantly affects the likelihood of conforming. Formally, we define the variable *Rate_others* for every minority player i as

$$Rate\_others(i) = \frac{\sum_{j,k \in N: j \neq i, \theta_j = down, \theta_k \neq down} \bar{g}_{jk}}{\sum_{j,k \in N: j,k \neq i, \theta_j = \theta_k = down} \bar{g}_{jk}} \tag{8}$$

In summary, our regressions show that the probability for a minority player to conform is sensitive to both (1) his own linking activity, and (2) the linking activity of other minority players. This observation provides support for our signaling hypothesis.

4.2 Exogenous almost complete network

We turn now to a more direct examination of the role of linking as a communication device. The strategy here is to take dense networks that were created by subjects in the treatment **ENDO** and set them up as exogenous networks and have the subjects play coordination games on these networks. The thought here is that if links do not play a communication role, then they most likely are primarily effective by altering payoff trade-offs. And if that is the case then behavior in the endogenous linking treatment must be similar to the

	1	2	3	4
SUBS	-0.50	-0.43	-0.50	-0.27
	(0.62)	(0.61)	(0.56)	(0.57)
COST	0.71*	0.54	0.89**	0.32
	(0.37)	(0.41)	(0.41)	(0.46)
period	-0.04***	-0.04***	0.04***	0.04***
	(0.01)	(0.01)	(0.01)	(0.01)
Rate	0.77***	0.88***	0.71***	0.67 ***
	(0.25)	(0.28)	(0.24)	(0.25)
Degree		-0.03	-0.04	-0.02
		(0.03)	(0.03)	(0.03)
Rate_others			0.35***	0.62***
			(0.12)	(0.14)
Degree_others				-0.22***
				(0.07)
SUBS×Rate	-0.22	-0.20	-0.10	-0.21
	(0.42)	(0.40)	(0.35)	(0.37)
COST×Rate	-0.57**	-0.66**	-0.45*	-0.49*
	(0.26)	(0.28)	(0.24)	(0.25)
Constant	-2.02***	-1.68***	-2.27***	-1.07*
	(0.34)	(0.47)	(0.51)	(0.62)
N	2,520	2,520	2.520	2520

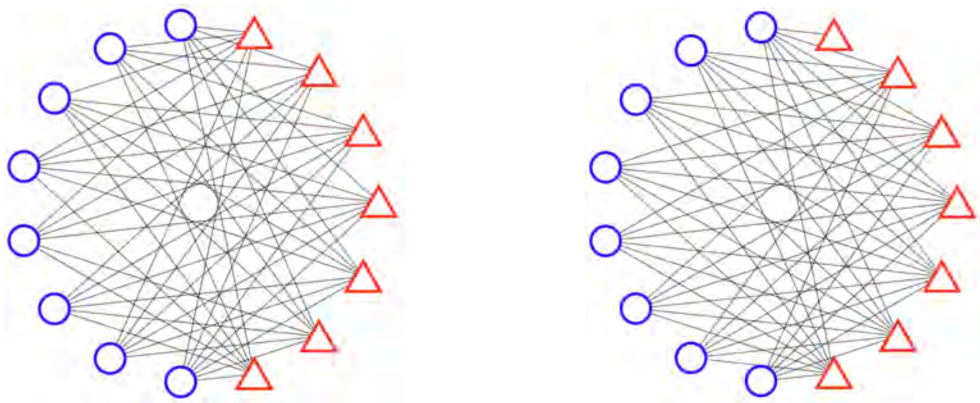
Confidence levels at 99%(***), 95%(**), and 90%(*).

Table 2: Probit Regression of *conformity* pooling all treatments with endogenous networks (**ENDO**, **SUBS**, and **COST**).

behavior in the (corresponding) exogenous networks.

There are a range of networks observed in the **ENDO** treatment: we take two distinct network configurations with the same number of missing links (i.e., 7), leading to a 87.5% connectivity across types (there is fully connectivity within the same type).[10] For robustness, we consider one symmetric and one asymmetric pattern of missing links. In **EXOSYM** every minority player has exactly one missing link (see Figure 12(a)). In **EXOASYM** 1 minority player is missing all links with majority players while the remaining 6 minority players are connected to all the majority players (see Figure 12(b) where

[10] Figure 7 revealed that across types linking ratio oscillates between 80% and 90% in **ENDO**.



(a) **EXOSYM** (b) **EXOASYM**

Figure 12: Across types connectivity in exogenous incomplete networks

the top triangle player is least connected with the majority).[11]

We present the equilibrium analysis of the coordination game in these networks.

Proposition 7. *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 if $n \geq \alpha/\beta + |N_d|$.

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

The proof is presented in Appendix A. The main point to note is that conformity (on up or down) and diversity both remain equilibrium outcomes under **EXOSYM** and **EXOASYM**. Figure 12 illustrates the diversity outcomes.

There were 6 groups for each of the two network treatments. Under **EXOSYM**, one group converges to conformity on the majority's preferred action, one group converges to conformity on the minority's preferred action, and the remaining four groups converge to diversity. Under **EXOASYM**, three groups converge to conformity on the majority's preferred action and the remaining three groups converge to diversity. As a result, while

[11]Moreover, in both **EXOSYM** and **EXOASYM**, we ensure that no majority player has more than one missing link with a minority player. This is a simplifying assumption as the majority player's behavior does not vary between **EXO** and **ENDO**.

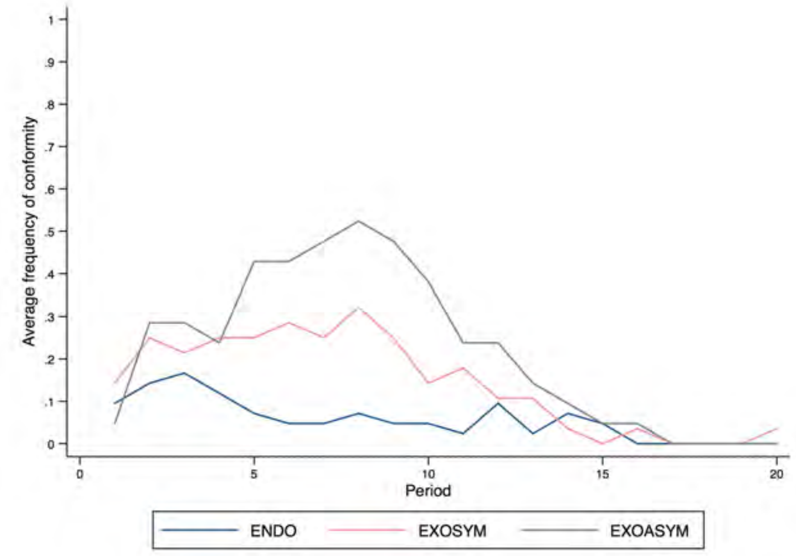


Figure 13: Minority action choice in **EXOSYM** and **EXOASYM** compared to **ENDO** (groups converging to diversity only).

the diversity outcome was reached in all 6 groups (i.e., in 100% of the cases) under **ENDO**, it was attained in only 7 out of 12 groups (58%) in under **EXOSYM** and **EXOASYM**. To summarize:

Experimental Finding 4. *In the almost complete exogenous networks, subjects choose conformity in over* 40% *of the cases. By contrast, when the same networks are endogenously created, subjects choose diversity in 100% of the cases.*

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

In order to further bring out this role of linking, we depict average behavior of the minority players within the groups converging to diversity in Figure 13. Moreover, convergence to diversity is much slower under **EXOASYM** and under **EXOSYM** relative to **ENDO**. In our view this offers support for the signaling role of links: the absence of linking in **EXOASYM** and **EXOSYM** prevents players from signalling their intentions and this slows down convergence in actions.

5 Alternative 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.

5.1 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.[12]

First consider an exogenous complete network g. In any period $t > 1$, the dynamic process is described as follows. In each period, 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 period, 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 $1-\epsilon$ 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 periods. 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 period 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 period (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

[12]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].

experiment, Proposition 8 specifies the existence of a unique stochastically stable state in the **EXO** treatment.

Proposition 8. *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. According to Proposition 8, 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 period t and $s^t = (g^t, x^t)$ denote the action profile at the end of period t (where x^t is as in the exogenous case previously described). In any arbitrary period 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 period 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 9. *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. 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 8). This result is clearly inconsistent with behavior observed in the **ENDO** and **SUBS** treatments. Thus, stochastic stability cannot provides an adequate explanation for the behavioral patterns observed in our experiment.

5.2 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]. It is however not obvious what labeling cue(s) could be exploited in our experiment setting. 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.[13] 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 (TR) 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.[14] In a complete network, this assumption of same-type similarity in behavior considerably simplifies the decision problem, as summarized in the following result.

Proposition 10. *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 TR equilibrium.*

The proof is presented in Appendix A. Our empirical observations from **EXO** are consistent with Proposition 10. In particular, the above equilibrium is justified as follows: it is strictly dominant for the majority to play their preferred action, and knowing this,

[13] As an example, the collective payoff of a group can be determined as the average individual payoff among its members.

[14] 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.

the minority is better off conforming to the majority's preferred action (see the proof of Proposition 10 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 11. *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 player with the majority ($0 \leq m_d \leq |N_d|$). If $|N_u| > |N_d|$, $\frac{|N_d|}{|N_u|} < \frac{\alpha-\beta}{\beta}$, and $\frac{|N_d|}{|N_u|-1} \geq \frac{\beta}{\alpha-\beta}$, then a TR 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 $k \leq 0$.*

The proof is presented in Appendix A. The condition of Proposition 11 is consistent with all our experimental treatments involving endogenous linking. In the baseline scenario where $k = 0$, it is clear that the minority's linking activity in the first stage plays an important signalling role for their subsequent behavior in the second stage. More specifically, all minority players forming links with the majority 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 signals their intention to select their preferred action afterwards. While our observations in **ENDO** are consistent with this kind of equilibrium behavior (according to Figure 7, 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 TR equilibrium, i.e., why did subjects select partial integration with diversity rather than full integration with conformity?

In previous studies, it has been 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 11). 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.

5.3 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 conformism 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).

5.4 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.

6 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 the three equilibria described above, or the network is only partially connected. In the latter case the network may fragment into two distinct 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 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 a choice, subjects form dense networks but choose diverse actions. The freedom to link with others allows individuals to differentiate themselves and this greatly reinforces diversity. Standard theories – such as stochastic stability, team reasoning, k-level reasoning, inequity aversion – cannot account for this pattern of behavior. We suggest that players use linking to communicate the action they will play in the coordination game.

References

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

G. Akerlof and R. Kranton. Identity and economics. Quarterly Journal of Economics, 115: 1–39, 2000.

A. Alesina, Baqir, and W. Easterly. Public goods and ethnic divisions. Quarterly Journal of Economics, 114:1243–1284, 1999.

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.

A. Bisin and T. Verdier. Beyond the melting pot: Cultural transmission, marriage, and the evolution of ethnic and religious traits. Quarterly Journal of Economics, 115:955–988, 2000.

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. Melndez-Jimnez, and M. Sutter. Experimental games on networks: Underpinnings of behavior and equilibrium selection. Econometrica, 82:1615–1670, 2014.

R. Chen and Y. Chen. The potential of social identity for equilibrium selection. American Economic Review, 101:2562–2589, 2011.

V. Crawford and B. Broseta. What price coordination? the efficiency-enhancing effect of auctioning the right to play. American Economic Review, pages 198–225, 1998.

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.

F. Fukuyama. Identity, immigration and liberal democracy. Journal of Democracy, 17(2): 5–20, 2006.

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, and Y. Vorobeychik. Behavioral experiments on a network formation game. ACM EC 2012, 2012.

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

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.

T. C. Schelling. Micromotives and Macrobehavior. Norton, 1978.

R. Sethi and R. Somanathan. Inequality and segregation. Journal of Political Economy, pages 1296–1321, 2004.

M. Sherif, O. Harvey, B. White, W. Hood, and C. Sherif. The Robbers Cave Experiment: Intergroup Conflict and Cooperation. Wesleyan University Press, 1988.

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

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

J. B. van Huyk, R. Battalio, and R. Beil. Strategic uncertainty, equilibrium selection, and coordination failure in average opinion games. Quaterly Journal of Economics, 106: 885–910, 1991.

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

Appendix A Proofs

Proof of Proposition 4:

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

$$W(x, y) = (n - x - y)(\alpha(|N_u| - x) + \beta(|N_d| - y)) + (x + y)(\beta x + \alpha y). \tag{9}$$

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

$$x^* = \frac{\beta(|N_d| - 2y) + \alpha(|N_u| - 2y) + \alpha(n)}{2(\alpha + \beta)}. \tag{10}$$

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

$$y^* = \frac{\alpha(|N_u| - 2x) + \beta(|N_d| - 2x) + \beta(n)}{2(\alpha + \beta)} \tag{11}$$

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

$$\frac{|N_u|}{|N_d|} > \frac{\alpha - \beta}{\alpha + \beta} \tag{12}$$

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

$$\frac{|N_d|}{|N_u|} > \frac{\alpha - \beta}{\alpha + \beta} \tag{13}$$

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 conformism 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 conformism on the majority action. This completes the proof. □

Proof of Proposition 7:

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. □

Proof of Proposition 8:

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 characterised by the set of Nash equilibria in pure strategies as specified by Proposition 2. 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 9:

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 8 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 9 then directly follows from the proof of Proposition 8. □

Proof of Proposition 10:

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 11:

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. □

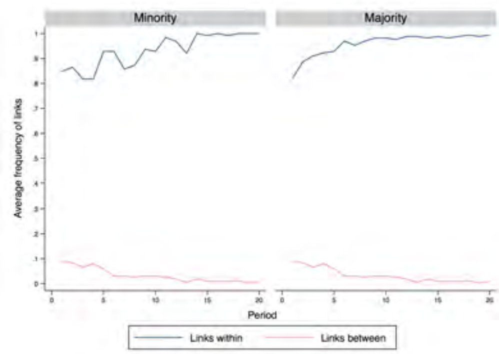
Appendix B Additional Experiments

This section presents two additional experiments we ran: **COST+** and **SUBS-**. The former 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 $\alpha = 6$, $\beta = 4$. The latter is a treatment with lower subsidy compared to **SUBS**, given by $k = -0.1 > -0.3$. the game is as in **SUBS** so that the value of other parameters remain $\alpha = 4$ and $\beta = 2$.

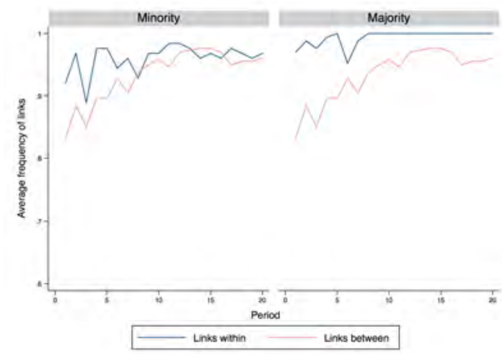
First, we look at the main findings in the **COST+** treatment. Consistent with the results from **COST**, we observe in Figure 14a that increasing the cost of linking leads to lower linking across preference types compared to **COST**. (Wilcoxon-Mann Whitney: $z = 7.142$, $p < 0.0001$), but that it does not significantly affect the long run outcome: we still observe (almost complete) segregation and diversity (see Figure 15-left). Notably, in **COST+** individuals choose actions more or less in line with their preferences. The majority chooses its preferred action almost from the start and persist with it for the entire experiment. The minority players mostly choose their preferred action during the first rounds of play and by period 11 no minority player is choosing the action of the majority. The effect of linking costs is clear, segregation and diversity.

For the second treatment, **SUBS-**, consistent with the results in **SUBS**, a positive subsidy for linking (negative cost) increases the level of connectivity. Moreover, there are no significant differences in network density between treatments (Wilcoxon-Mann Whitney: $z = 1.603$, $p = 0.1089$). This is particularly clear when looking at the high level of between group connectivity in Figure 14b, which as stated before, is not different from **SUBS** (Wilcoxon-Mann Whitney: $z = 0.202$, $p = 0.8398$). Diversity of actions is a prominent outcome in **SUBS-** with 50% of the groups portraying complete diversity (see Figure 15-right)[15]. Thus, a positive subsidy promotes integration and to a large extent diversity in actions.

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



(a) COST+



(b) SUBS-

Figure 14: Linking choices in **COST+** (left) and **SUBS-** (right).

To conclude this section, we extend the regressions reported in Table 2 by considering the data from all treatments including **COST+** and **SUBS-**. The new regressions, reported in Table 3 indicate, that the linking ratio is the strongest predictor of conformity. Thus, providing support to the findings presented before. Notice, however, that given there are some cases of conformity in **SUBS-** this treatment has a positive and significant coefficient in Model (1). Moreover, we observe a small but significant effect of individual degree, thus suggesting that the effect of the linking ratio may be moderated by a player's own degree (see Model 2). Additionally, the ratio of other minority players has a positive effect on conformity, as shown in Model (3). Note however that such effects of *Degree* and *Rate_others* are limited compared to that of *Rate*. In conclusion, there is a consistent and dominant effect of endogenous linking on diversity (i.e. no conformity).

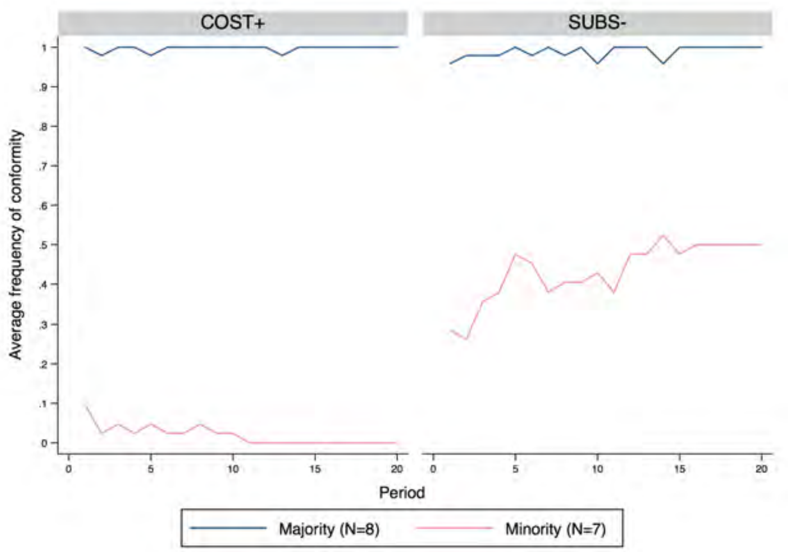


Figure 15: Conformity in **COST+** and **SUBS-**.

	1	2	3	4
SUBS	-0.43	-0.18	-0.34	-0.31
	(0.62)	(0.61)	(0.55)	(0.55)
SUBS-	0.96**	0.97**	0.91	0.92*
	(0.44)	(0.48)	(0.42)	(0.42)
COST	0.78**	0.64	1.40***	0.97**
	(0.38)	(0.44)	(0.41)	(0.43)
COST+	0.29	0.27	1.28***	0.80*
	(0.42)	(0.47)	(0.46)	(0.48)
Period	-0.01***	-0.01**	-0.01***	-0.01**
	(0.004)	(0.004)	(0.004)	(0.004)
Rate	0.75***	0.94***	0.64***	0.60***
	(0.23)	(0.27)	(0.21)	(0.21)
Degree		-0.05**	-0.04**	-0.03
		(0.02)	(0.02)	(0.02)
Rate_others			0.57***	0.66***
			(0.09)	(0.10)
Degree_others				-0.11**
				(0.05)
SUBS×Rate	-0.20	-0.29	-0.13	-0.13
	(0.39)	(0.37)	(0.33)	(0.33)
SUBS-×Rate	0.21	0.18	0.22	0.22
	(0.30)	(0.32)	(0.28)	(0.27)
COST×Rate	-0.61**	-0.72***	-0.56**	-0.49**
	(0.24)	(0.27)	(0.22)	(0.21)
COST+×Rate	-0.05	-0.25	0.01	0.02
	(0.33)	(0.35)	(0.30)	(0.29)
Constant	-2.24***	-1.84***	-2.96***	-2.30***
	(0.35)	(0.45)	(0.46)	(0.51)
N	4,200	4,200	4,200	4,200

Confidence levels at 99%(***), 95%(**), and 90%(*).

Table 3: Probit Regression of *conformity* Pooling all treatments with endogenous networks (**ENDO**, **SUBS**, **COST**, **SUBS-**, and **COST+**).

Appendix C 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 participants take place through the computers. Please, do not talk to the other participants 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 participants 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 participants 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 participant 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.

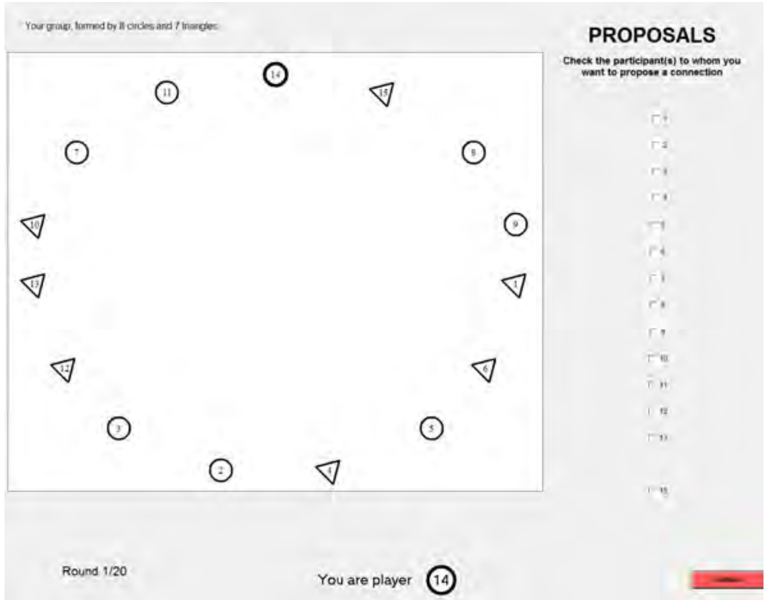
Specific to Treatment ENDO only:

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 participants, 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 participants 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.



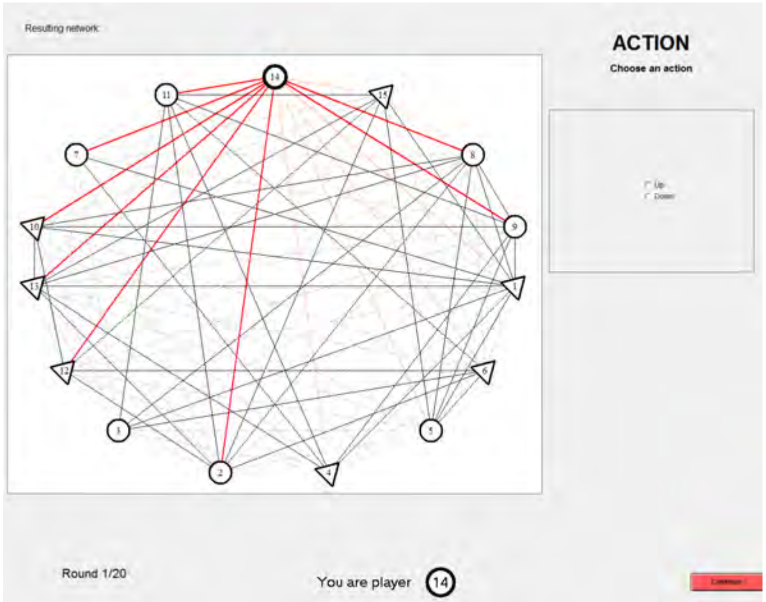
A connection is formed if 2 participants propose to each other. In Phase 2 (Action) you will interact only with the participants to whom are connected.

Phase 2. Action

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

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

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 participants 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:

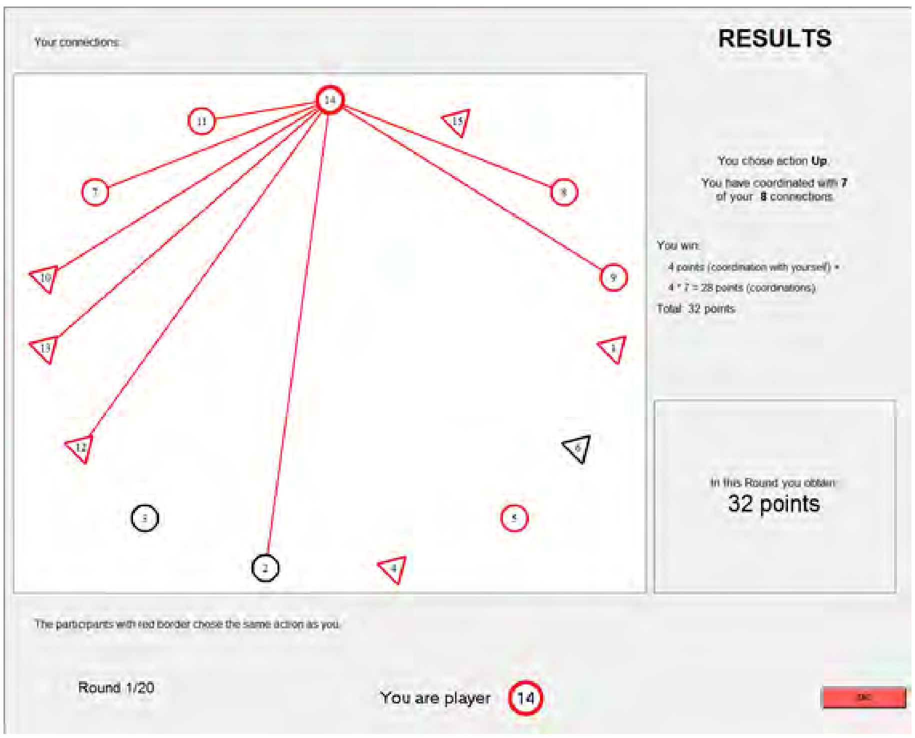
- choose **down**, you receive **4 points for each** of your connections choosing **down**
- choose **up**, you receive **2 points for each** of your connections choosing **up**

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 participants 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 participant 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 participants, 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 participants, 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.

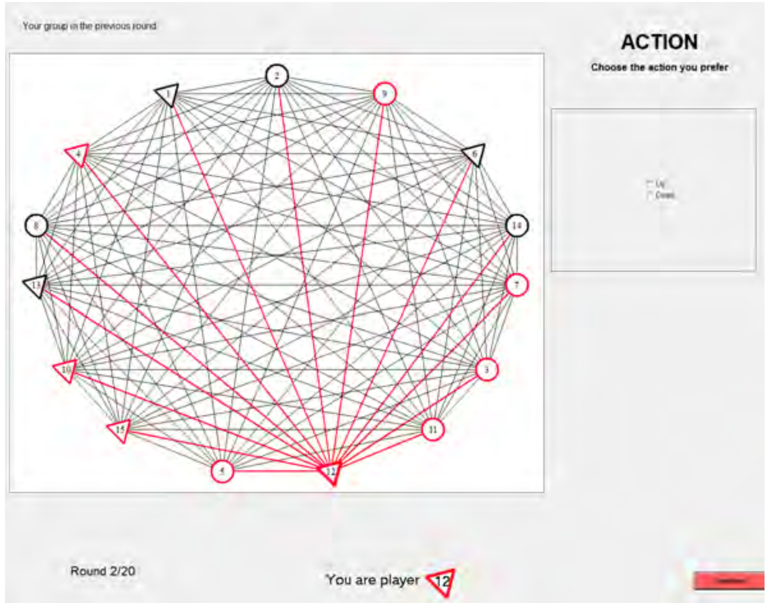
Specific to Treatment EXO only:

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 participants you interact with and their choices in the previous round (in the first round you will see the participants without any previous decision). You will see your number and your type, and the numbers and types of the other participants, 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 participants 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:

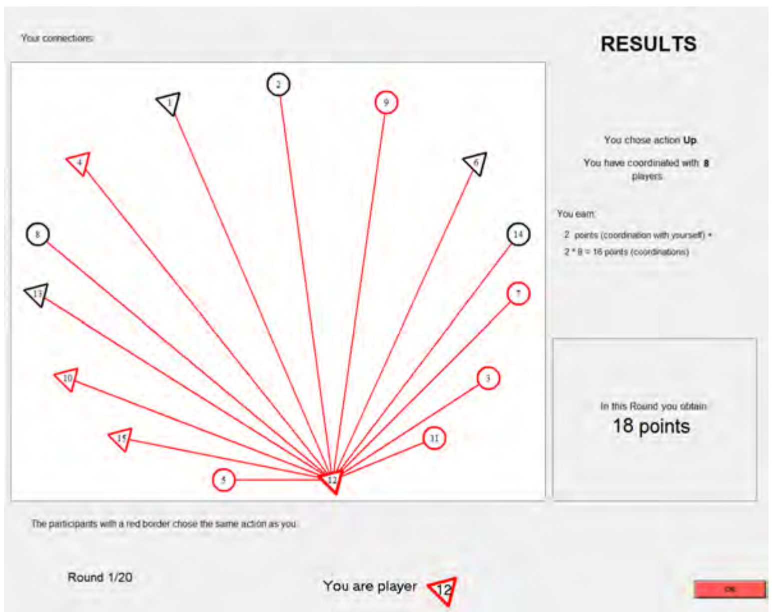
- choose **down**, you receive **4 points for each** of your connections choosing **down**
- choose **up**, you receive **2 points for each** of your connections choosing **up**

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 participants 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 participant 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, me present two examples:

Example 1: You are a circle, you have chosen up and 4 other participants 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 participants 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 participants 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 participants for a whole stage (trial or experiment), but your group, symbol and number, and those of the other participants, might change between stages.