Integration and Segregation

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

May 24, 2017

Abstract

Individuals prefer to coordinate with others, but they differ on the preferred action. In theory, this can give rise to an integrated society with everyone conforming to the same action or a segregated society with members of different groups choosing diverse actions. Social welfare is maximum when society is integrated and everyone conforms on the majority's action. In laboratory experiments, subjects with different preferences segregate into distinct groups and choose diverse actions. To understand the role of partner choice, we then consider an exogenous network of partners. Subjects in the experiment now choose to conform on the action preferred by the majority. Thus, there exists a tension between two deeply held values: social cohesion and freedom of association.

[*]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
[¶]Departamento de Análisis Económico, Universitat de València, Email: edson.munoz@uv.es
[‖]Grupo Interdisciplinar de Sistemas Complejos, Departamento de Matemáticas, and 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 Vince Crawford, Matt Elliott, Edoardo Gallo, Jonathan Newton, Theo Offerman, Gustavo Paez, Arno Reidl, Robert Sugden and participants at a number of seminars for helpful comments.

1 Introduction

We study the relation between two deeply held values: social cohesion and freedom of association. Social cohesion calls for wide interpersonal interaction and is supported by shared norms of behavior. The interest in social cohesion is not limited to academic discourse: threats to shared norms and values are a prominent topic of discussion in contemporary American and European politics. Differences in preferences across modes of behavior is a recurring theme in these debates.[1] Freedom of association includes the right to form relationships to join or leave groups of a person's own choosing.[2] This paper examines the role of freedom of association in sustaining shared norms of behavior, when individuals have differing views on preferred norms.

To clarify the key considerations and in order to be able to discuss social welfare, we start by setting out a theoretical model. There is a group of individuals who each choose between two actions a and b. Everyone prefers to coordinate on one action but individuals differ in the action they prefer: group A prefers action a, group B prefers action b. We consider a baseline setting in which everyone is obliged to interact with everyone else. And a setting that reflects freedom of association: individuals decide with whom they interact. In the latter setting, everyone observes the network that is created and then chooses between action a and b. The theoretical analysis reveals a rich set of possibilities.

Consider first the case where everyone interacts with everyone else. There exist three equilibria: everyone conforming to action a, everyone conforming to action b, and diversity (with group A members choosing a and group B choosing b). The diversity outcome is sustainable only when neither of the groups is too small. Aggregate welfare is defined as the sum of individual payoffs. We show that conforming to the majority action maximizes social welfare.

Consider next the setting with freedom of association. Again, there exist multiple possible stable outcomes: broadly, they take on two forms. In one case, everyone connects with everyone else and chooses the same action. We refer to this as integration with conformism. The other case exhibits separation of the community into two distinct compo-

[1] For example, in January 2017, the Dutch Prime Minister brought out a Public Advertisement in which he urged individuals belonging to minority communities to 'be normal' or 'to leave the country'. A key aspect of the Dutch context is that individuals differ in their preferences over norms. In the Dutch context, a prominent case pertains to a male applicant for the job of a driver at a coach company who did not want to shake hands with female passengers. The company refused to hire him.

[2] It is guaranteed by the United States Bill of Rights, the European Convention on Human Rights, and the Universal Declaration of Human Rights.

nents: everyone within a component chooses the same action. Integration and conforming on the majority action is welfare maximizing.

Thus, the theory yields a clear cut prediction on welfare: in both the exogenous and endogenous interaction setting, conforming to the majority action maximizes aggregate welfare. However, a variety of other – inefficient – outcomes are possible under both the exogenous and the endogenous linking setting. So, to get a better sense of the role of freedom of association, we turn to laboratory experiments.

The experiments involve groups of 15 subjects, who play a coordination game repeatedly over 20 rounds. There is a majority sub-group with 8 subjects and a minority sub-group with 7 subjects. The principal finding is that, with exogenous interaction, conformity on the majority obtains; by contrast, in the endogenous interaction setting there is almost complete segregation (between the majority and minority) and a diversity of actions. So our experiment suggests that when individuals differ on preferred modes of behavior, there exists a tension between the principle of freedom of association, on the one hand, and social cohesion and welfare, on the other hand. This tension creates large welfare losses. Figures 7 and 8 illustrate the dynamics leading to these outcomes.

A closer examination of the dynamics of the linking and action choice in the experiment reveals that in the exogenous network setting, in round 1, around 40% of minority conforms to the majority action. Gradually, over time this fraction grows and there is close to full conformity by round 15. By contrast, in the endogenous linking treatments, in round 1, less than 20% of the minority choose the majority action. This suggests that differences in outcomes at round 1 are sharp and dramatic. Indeed, given these initial conditions, minority individuals who choose a myopic best response would be led to conformism on the majority action in the exogenous setting, and will remain segregated and choose their own preferred action in the endogenous setting.

We then turn to a theoretical explanation for observed behavior. We explore four approaches: stochastic stability, team reasoning, social preferences, and level-k reasoning. The predictions of the former two approaches are more clear cut and we summarize them here: both stochastic stability and team reasoning predict conformism on majority action with exogenous networks. The predictions under endogenous linking are more varied: both integration and segregation are dynamically stable under endogenous linking. Similarly, under team reasoning, segregation and diversity are sustainable (along with other outcomes) under endogenous networks. Thus the experimental finding of conformism on majority action is strongly consistent with the two theories, while the finding of segregation and diversity is only ‘weakly’ consistent with the theory.

Our paper is a contribution to the study of social coordination. Experiments on coordination have been a very active field of research, Charness et al. [2014], Crawford [1995], Isoni et al. [2014], Kearns et al. [2012]. In this literature, the minimum effort game has been especially prominent. 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 theoretical simplicity has motivated a number of experimental investigations of this game, over the years. The early experiments 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 [1991], Crawford and Broseta [1998]. Our paper is closely 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.[3] A key factor explaining their results is that players are ostracized (and isolated) if they do not contribute the maximum effort. However, if they increase their contribution (i.e. behave as expected by the group), then they are welcomed back in the group. This leads eventually to a completely connected network contributing the maximum effort level. In the present paper, individuals have differing preferences across social outcomes. Our experiment reveals that these differences become salient when linking is endogenous. Thus, our work shows that endogenizing linking can have very different consequences, depending on whether or not individuals have heterogeneous or similar preferences.

We turn finally to the experiments on identity and coordination games. Types in our setting may naturally be interpreted as an aspect of identity.[4] In particular, following the

[3]More generally, following on the work of Blume [1993] and Ellison [1993], interest has centered on the role of interaction structures in shaping coordination. 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 recent work, Advani and Reich [2015], Bojanowski and Buskens [2011] and Ellwardt et al. [2016] introduce heterogeneous preferences in the framework with partner choice and action choice.
There is also a strand of work on the role of endogenous interactions in cooperation, see e.g., Wang et al. [2012], Gallo and Yan [2015], Cuesta et al. [2015]. For a study of the effects of endogenous groups in public goods games, see Kosfeld et al. [2009].

[4]There is a large literature on identity, spanning across several disciplines in the social sciences and in philosophy. For an early and influential study in economics of how individual choices may lead to segregation and the the break down of social cohesion, see Schelling [1978]. In recent years, there has been a great deal of interest in understanding the ways in which identity shapes behavior in society, organizations, markets,

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 the minimum effort game.

In our setting identities are reflected in payoff differences and they are kept constant. We interpret our findings as suggesting that freedom of association allows identity 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 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 action. In the experiment, subjects create 'almost' complete networks but different types nevertheless choose their preferred actions! This suggests that the very possibility of linking makes different preferences/identities salient. Figure 14 illustrates the dynamics in the zero cost setting.

The paper is organized as follows. Section 2 presents the model and a theoretical analysis. Section 3 presents our experimental design, procedure and the experimental findings. Section 4 takes up theoretical explanations. Section 5 concludes. Appendix A contains some of the proofs, while Appendix B contains the instructions for the experiments.

2 Theory

We study a network formation and action choice game 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 characterise the nature of equilibrium action choices in networks.

---

and in local government, see e.g., Sethi and Somanathan [2004], Akerlof and Kranton [2000], Alesina et al. [1999], Advani and Reich [2015], Bisin and Verdier [2000].

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 $A = \{a, b\}$. Formally, $x_i : \overline{G} \to A$, and we define X_i as the set of all such strategies for player i. We denote the set of overall strategies of player i in the full game as $S_i = G_i \times X_i$, and the set of overall strategies for all players as $S = S_1 \times \ldots \times S_n$. A strategy profile $s = (x, g)$ specifies the link proposals made by every player in the first stage through $g = (g_1, g_2, \ldots, g_n)$, and the choice functions made by each player in the second stage through $x = (x_1, x_2, \ldots, x_n)$. We define $N_i(\overline{g}) = \{j \in N : \overline{g}_{ij} = 1\}$ as the set of i's neighbours in the network $\overline{g}$.

Moreover, for every player i, let $\theta_i \in \{a, b\}$ define i's type. This leads us to define $N_a = \{i \in N : \theta_i = a\}$ and $N_b = \{i \in N : \theta_i = b\}$ as the groups of players preferring action a and b respectively ($N_a \cup N_b = N$). If $|N_a| \neq |N_b|$, we refer to the largest group of players sharing the same type/preferences as the *majority* and the other 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].[5]

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 some elementary 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

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

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.

[5] For related work on coordination games with heterogenous preferences in exogenous networks, see Hernandez et al. [2013] and Neary [2012].

Proposition 2. *Fix a complete network g. Suppose x^* be a Nash equilibrium. Then*

(i) every player selects the same action, i.e., $x_i^* = x_j^* \in \{a, b\}$ for all $i, j \in N$ if $n \geq \alpha/\beta$.

(ii) every player selects their preferred action, i.e., $x_i^* = \theta_i$, for all $i \in N$ if $|N_a|, |N_b| \geq \frac{\beta(n+1)}{\alpha+\beta}$.

We sketch the proof here. It follows from Proposition 1 that conforming on one action is an equilibrium if both the majority and minority individuals have no profitable deviation. 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 $n \geq \alpha/\beta$.

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 $y \in \{a, b\}$. This inequality can be rewritten as

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

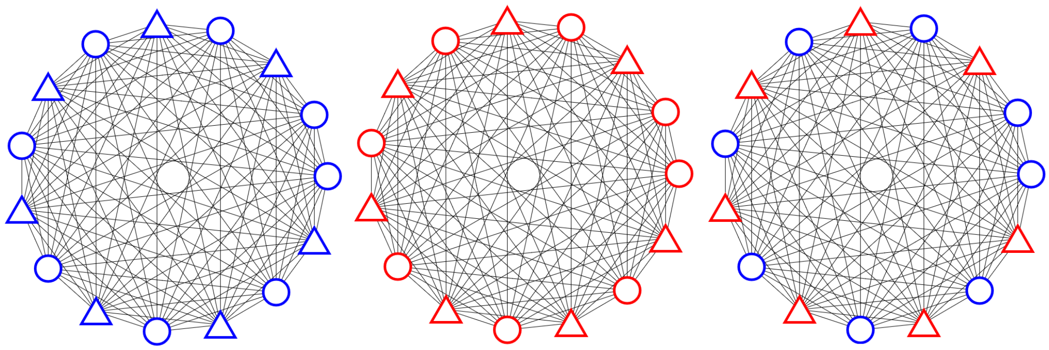
for any $y \in \{a, b\}$. This completes the argument.

So, we have shown that, in a complete network, there are only three equilibrium outcomes: *conformity* where every player coordinates on the same action, a or b, 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 fifteen individuals. There are 8 players in *circles* and the remaining 7 individuals are in *"triangles"*. The circles prefer action 'blue', while the triangles prefer 'red'.

Turning 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.

Proposition 3. *In a complete network, conformity on the majority's preferred action is socially efficient.*



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

Figure 1: Nash equilibria in complete network

Proof. Let x and y be the number of players playing b in N_a and N_b, respectively. The sum of individual payoffs is

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

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

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

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

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

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

$$\frac{|N_a|}{|N_b|} > \frac{\alpha - \beta}{\alpha + \beta} \tag{10}$$

This inequality holds whenever $|N_a| > |N_b|$.

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

$$\frac{|N_b|}{|N_a|} > \frac{\alpha - \beta}{\alpha + \beta} \tag{11}$$

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

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

□

The proof says that diversity is never socially desirable. The intuition is fairly simple: fixing the behavior of one group, it is never desirable for the other group to mix actions. This follows from the coordination externalities inherent in our model. So we only need to compare the two outcomes: one, where everyone conforms to action a and the other where everyone conforms to action b. The final step simply says that conformism on a is better if and only if the group that prefers a constitutes a majority. So, in our example, the socially efficient outcome corresponds to Figure 1(a).

We now analyze the link formation process in stage 1. 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 are 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, any two players who choose the same action in the second stage must also be linked with each other. On the other hand, in a pairwise stable network-action pair, any two players who choose different actions in the second stage are not linked to each other. *For any pairwise stable pair* $(\overline{g}, x(\overline{g}))$, $x_i(\overline{g}) = x_j(\overline{g})$ *if and only if* $\overline{g}_{ij} = 1$.

This means that in a pairwise stable network-action pair, there can be at most two components, each component is complete, and every individual in a component must choose the same action.

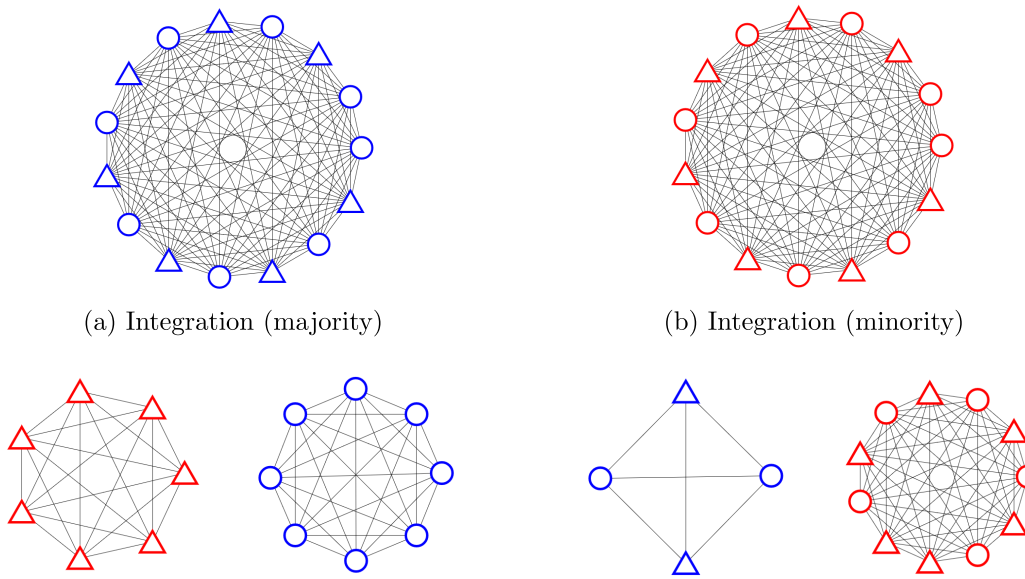
Proposition 4. *Suppose* $(\overline{g}^*, x^*(\overline{g}^*))$ *is pairwise stable. Then*

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

(ii) $\overline{g}^*$ *contains two complete components,* C_a *and* C_b*; every player in* C_a *chooses* a*, while every player in* C_b *chooses* b.

This allows us to define two types of outcomes: *integration with conformity*, where everyone links with everyone else, and chooses the same action and *segregation with diversity*, where individuals constitute two components, each of which chooses a different action.

The proof of the result is immediate, from the observations preceding it. In the integration with conformity outcome, every player forms links with all others, and they all coordinate on the same action, regardless of the type. The segregation with diversity outcomes do not depend on the players' types whenever the size of the smallest component is



(a) Integration (majority) (b) Integration (minority)

(c) Segregation by type (d) Mixed segregation

Figure 2: Pairwise stable outcomes

sufficiently large (otherwise, all players of the smallest group have to be of the same type and playing their preferred action).

It is worth drawing out the content of the result in the context of our example (with $n = 15$, $|N_{circle}| = 8$, and $|N_{triangle}| = 7$). The integration with conformism outcomes and the segregation with diversity outcomes are illustrated in Figure 2. The first part of the Figure represents the outcomes with integration and conformity and correspond to the outcomes under complete network. The second part of the Figure illustrates segregation outcomes.

We next note that with endogenous interaction, social welfare is maximized under integration and conformism on the majority's action.

Proposition 5. *In the game with linking and action choice, the socially efficient outcome always entails integration and conformity on the majority's preferred action.*

Proof. It is straightforward to see that in a socially efficient outcome, two players are linked with each other, if and only if, they choose the same action. This implies that it must be formed of at most two completely connected components. Let x and y be the number of players playing b in N_a and N_b respectively. We define the sum of individual payoffs in the corresponding outcome as

$$W(x,y) = (n-x-y)((\alpha-k)(|N_a|-x)+(\beta-k)(|N_b|-y))+(x+y)((\beta-k)x+(\alpha-k)y)+kn \tag{12}$$

If we replace α and β with $\alpha-k$ and $\beta-k$ in the proof of Proposition 3, then the proof follows. □

The integration equilibrium with conformity on the majority's preferred action is the unique socially efficient outcome whenever such a majority exists (i.e., $|N_a| \neq |N_b|$). In the context of our example, the socially efficient outcome corresponds to Figure 2(a).

To summarize: 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. Conformity is always welfare maximizing. In the endogenous setting, there exist multiple equilibria exhibiting integration with conformity and segregation with diversity. Finally, integration with conformity on the majority's preferred action maximizes welfare.

So diversity and conformism can arise under both exogenous as well as under endogenous interaction. Conformism on the majority action is always socially optimal. We conduct laboratory experiments to understand how freedom of association – interpreted here as choice of links – determines the level of integration, the extent of behavioral diversity, and the aggregate social welfare.

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 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 were informed of a symbol, either a circle or a triangle, and an identification number,

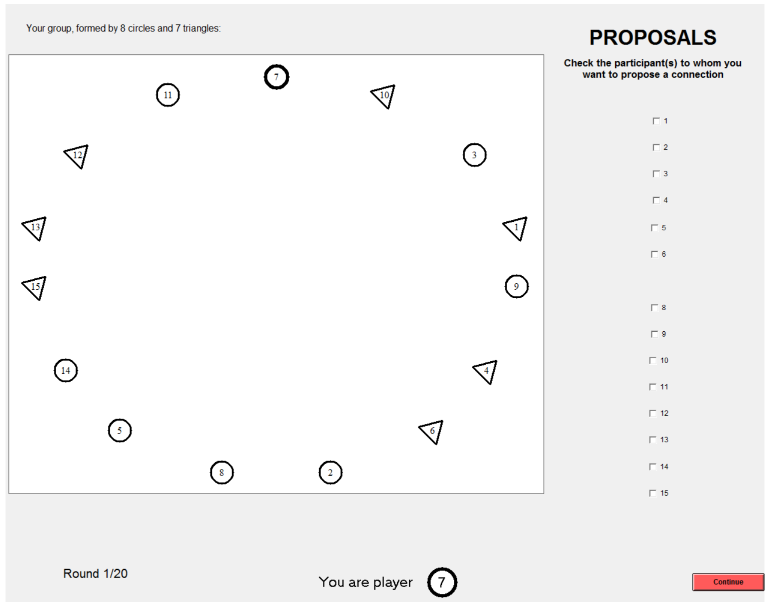


Figure 3: Individual Types

from 1 to 15, assigned to them. Every subject knew his symbol, number and the symbol and number of the 14 others in his group. Both symbol and number were 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 saw at the start of the experiment.

In the treatment **ENDO**, there were two stages. First subjects simultaneously made proposals to a subset of the others in their group. Reciprocated proposals led to the creation of links. If a link was formed then both subjects paid a cost of $k = 2$ points for it. Then, in the second stage, subjects were informed of the links proposed and those that were formed in stage 1. After observing the created network, subjects chose one of two actions: *up* or *down*. Figure 4 illustrates the network date that they observed at this point. In the screenshot, links that are proposed but not reciprocated are represented as light 'incomplete' links, while reciprocated proposals were represented as dark and 'complete' links. So in the screenshot, player 7 creates links with 4, 15, 8, 3 and 12. He does not reciprocate proposals from 9, 10, 14, 1, while he makes unreciprocated proposals to 11 and 2.

The values of the parameters are $\alpha = 6$, $\beta = 4$, and $k = 2$. For a subject with symbol

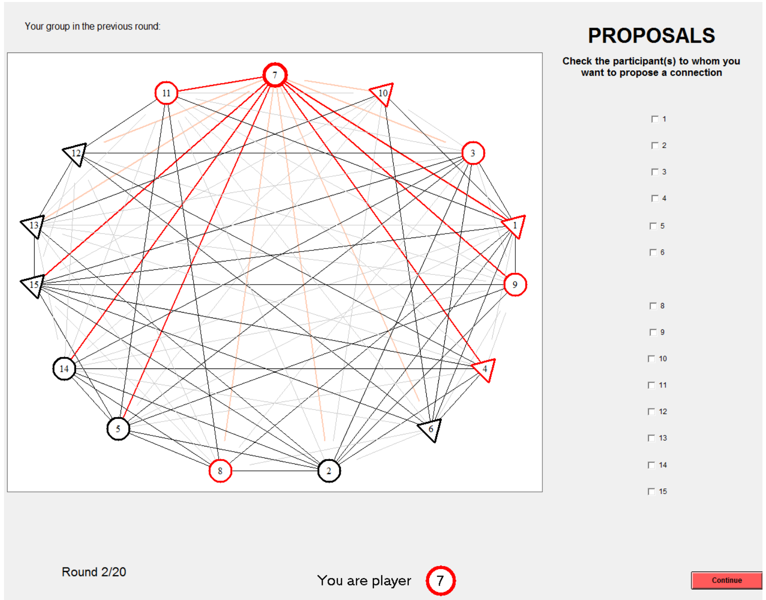


Figure 4: Linking proposals: reciprocated and unreciprocated

circle (triangle), his preferred action is up (down). Finally, every player sees the outcome of the game on the screen and his net payoffs. This screenshot is illustrated in Figure 5. Here we see that player 7's neighborhood includes 4, 15, 8, 3 and 12. He coordinates successfully on his less preferred action with players 15, 8 and 3, and he fails to coordinate with 4 and 12. Thus his net payoff is $4 \times 4 - 5 \times 2 = 6$. Finally, at the beginning of any round r>1, in stage 1, information about everyone's behavior in both stages of the previous round is provided to every subject, as shown through Figure 6.

In the treatment **EXO**, all subjects interacted with every other group member in a complete network. The subjects were shown the complete network and they had to choose between actions *up* and *down*.[6] Given that there is no linking decision, there are also no linking costs. For this reason, and to make earnings comparable between treatments, the parameters in **EXO** are: $\alpha = 4$, $\beta = 2$.[7] The Instructions handed out to subjects are presented in Appendix B.

[6]The complete network was however shown as it would be in **ENDO**, had the complete network emerged, see the instructions in the Appendix

[7]Being connected with a player who plays one's most preferred action is worth $\alpha = 4$ in **EXO** and $\alpha - k = 4$ in **ENDO**. Similarly, for the payoffs from the less preferred action, the payoff is 2 in both treatments.

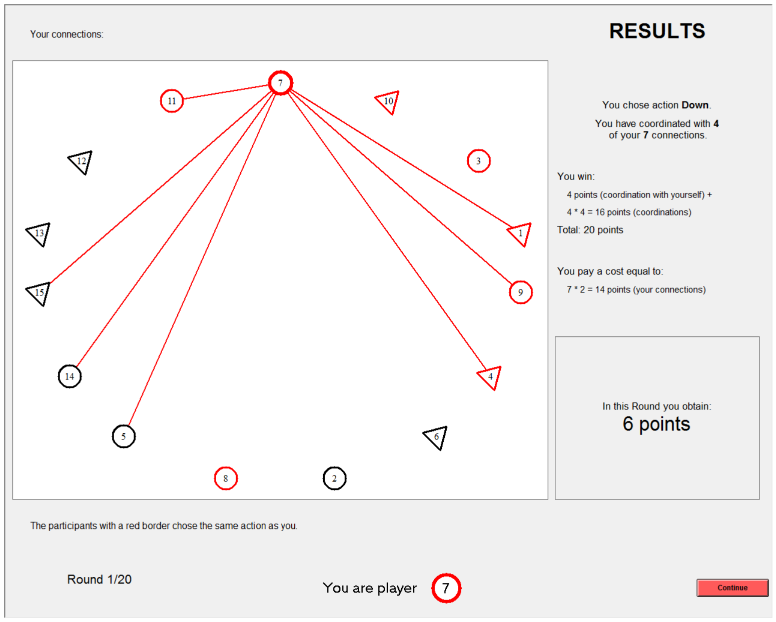


Figure 5: Neighborhood, Actions and Payoffs

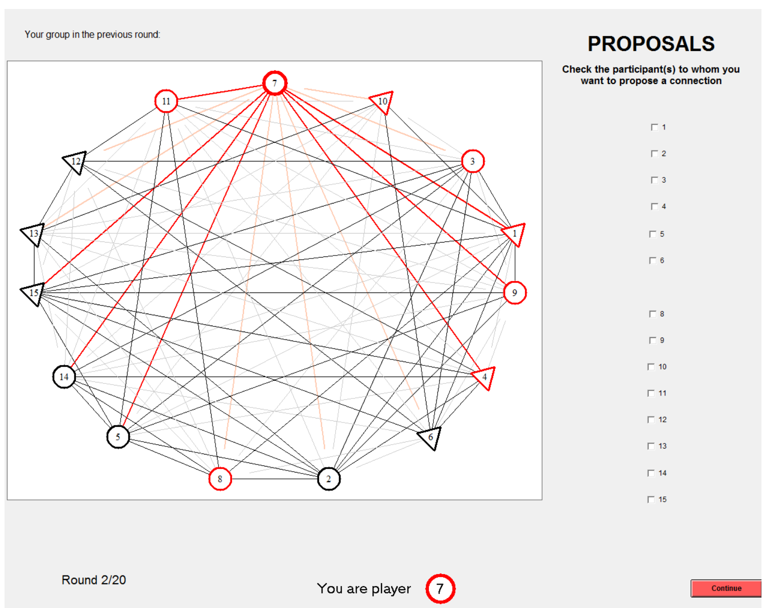


Figure 6: Feedback from previous round

To test the robustness of our main findings, we will consider variations on the two main treatments. These variations will entail different costs of linking and different relative sizes of the majority and minority groups. The following table provides a summary of all our treatments.

Distribution	**Network** ($N = 15$)			
	Endogenous (start from empty net)			**Exogenous** (complete net)
	$k = 2$, $\alpha = 6$, $\beta = 4$	$k = 0.5$, $\alpha = 4.5$, $\beta = 2.5$	$k = 0$, $\alpha = 4$, $\beta = 2$	-, $\alpha = 4$, $\beta = 2$
$\lvert N_O \rvert = 8$, $\lvert N_\Delta \rvert = 7$	**ENDO** (6 groups)	**CHEAP** (6 groups)	**FREE** (6 groups)	**EXO** (6 groups)
$\lvert N_O \rvert = 12$, $\lvert N_\Delta \rvert = 3$	**LOWCONFLICT** (6 groups)			

Table 1: Experimental Treatments

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. So there were 45 subjects per session. Each session lasted between 90 and 120 minutes, and on average subjects earned approximately 18 euros.

3.3 Experimental Findings

Throughout this section, 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 last five 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.

To get a first impression of the dynamics, Figure 7 depicts behavior observed for one group in treatment **EXO**. Similarly, Figure 8 depicts the dynamics for one group in treatment **ENDO**. In those figures, *blue* represents the circle players' preferred action and *red* represents the triangle players' preferred action. For simplicity, unreciprocated proposals are not shown in both figures. These figures point to a dramatic difference in action choice in the two treatments. Moreover, the different action choices are driven by very different patterns of linking. In the exogenous networks setting, we observe conformity on majority action, while in the endogenous networks setting we observe segregation and diversity.

These figures bring out the dramatic effect of endogenous linking on network architecture and on behavior clearly.

We now turn to a closer examination of the data on behavior in the coordination game, across the two treatments, **ENDO** and **EXO**. In **EXO**, most subjects choose the majority action, but in **ENDO** individuals choose actions more or less in line with their preferences. In particular, the average number of subjects choosing the majority's action across rounds is significantly lower in **ENDO** than in **EXO**, $8 < 14.68$ (Wilcoxon-Mann-Whitney: z=–4.34, p=0.00). Figure 9 shows that the difference between treatments is due to the choices by the minority. Indeed, there is practically no change in the behavior of the majority across these two treatments. The members of the minority, on the other hand, rarely choose the majority action under **ENDO** while they eventually converge to all choosing that action under the treatment **EXO** (z=-2.99, p=0.003).[8] In **ENDO** there are no players from the minority choosing majority action, while in **EXO** it cannot be distinguished from 7 (t=-1.73, p=0.16). To summarize: the majority chooses its preferred action almost from

[8]We note that 5 out of 6 groups in **EXO** achieve full conformity, in the 6th group the minority and the majority choose different behavior. 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 the group from the analysis and the illustration in Figure 9.

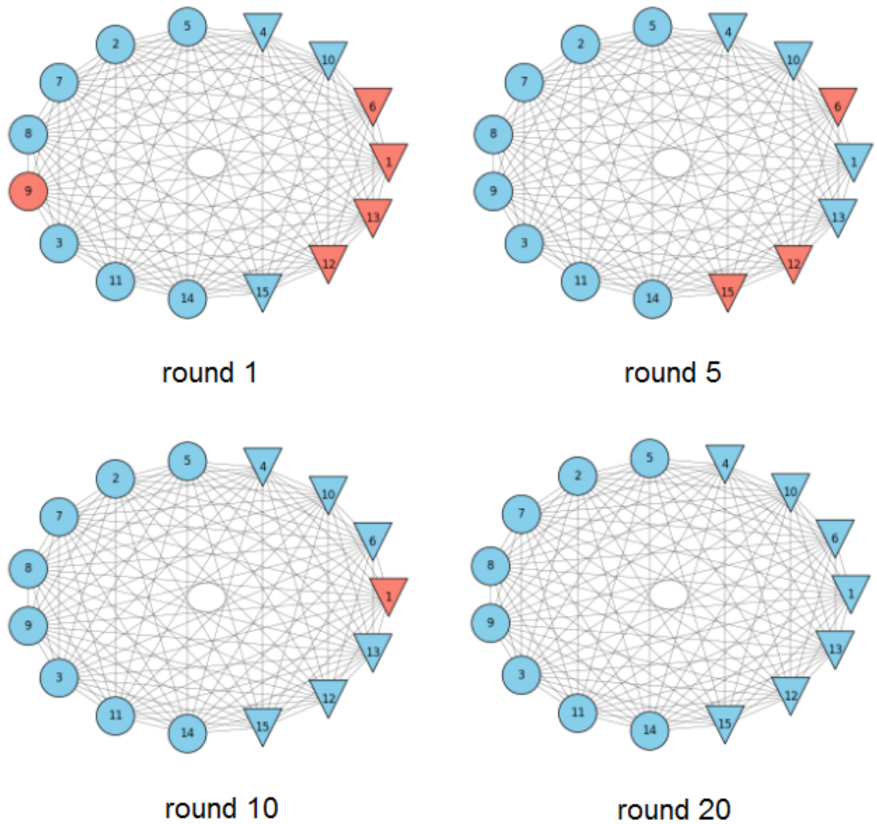


Figure 7: Illustration of dynamic behavior in **EXO**

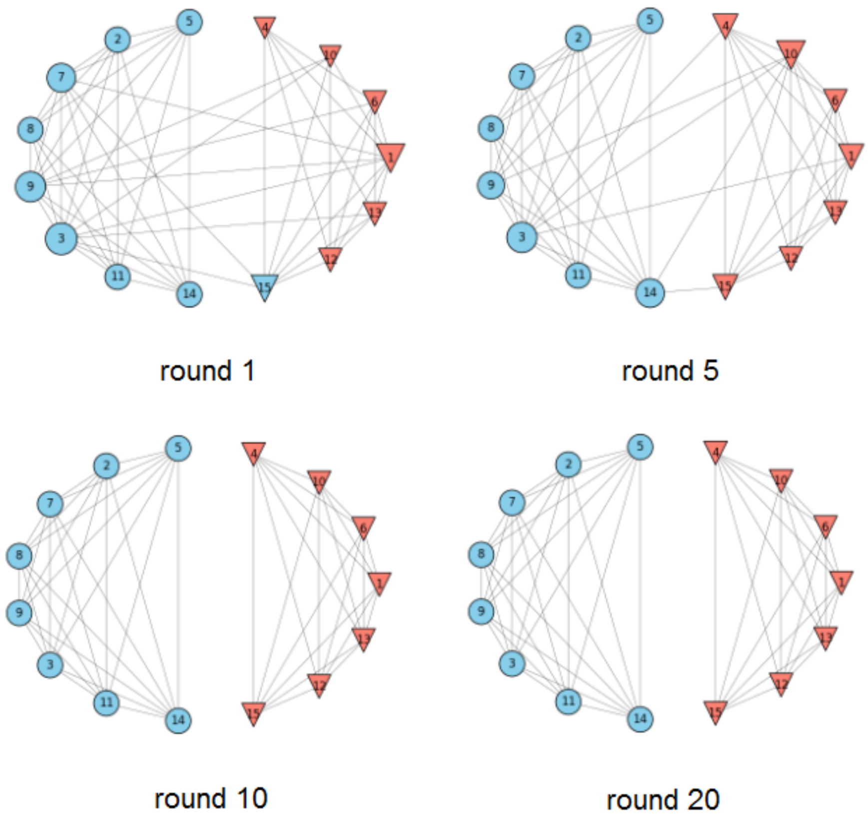


Figure 8: Illustration of dynamic behavior in **ENDO**

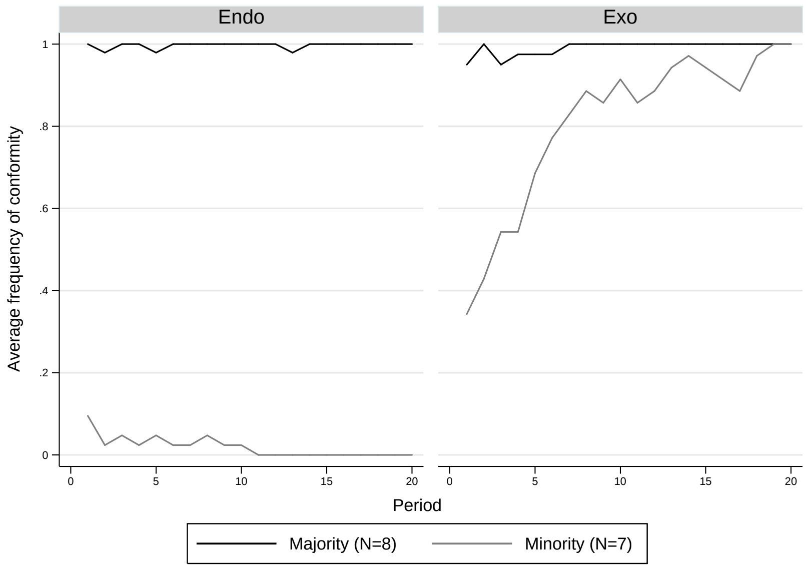


Figure 9: Action choice in **ENDO** and **EXO**.

the start under both treatments and persists with it throughout the rounds. The minority chooses its preferred action under **ENDO** and distinguishes itself almost fully from the majority under **ENDO**, from the start and persists with this choice. On the other hand, 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%. Everyone in the minority group is choosing the majority action by the end of the rounds and gradually this fraction converges to 100%.

We next discuss the data on linking choices in **ENDO**. At the start, almost 80% of the links within the groups are formed, but less than 20% of the cross group links are formed. By round 10, this tendency accentuates and over 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. Thus we see the emergence of complete segregation. This is illustrated in Figure 10. The average degree of the minority is 5.99 and that of the majority is 6.93.[9] Our first set of findings leads to:

Result 1. *With endogenous linking, subjects converge rapidly to segregation and diversity*

[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.

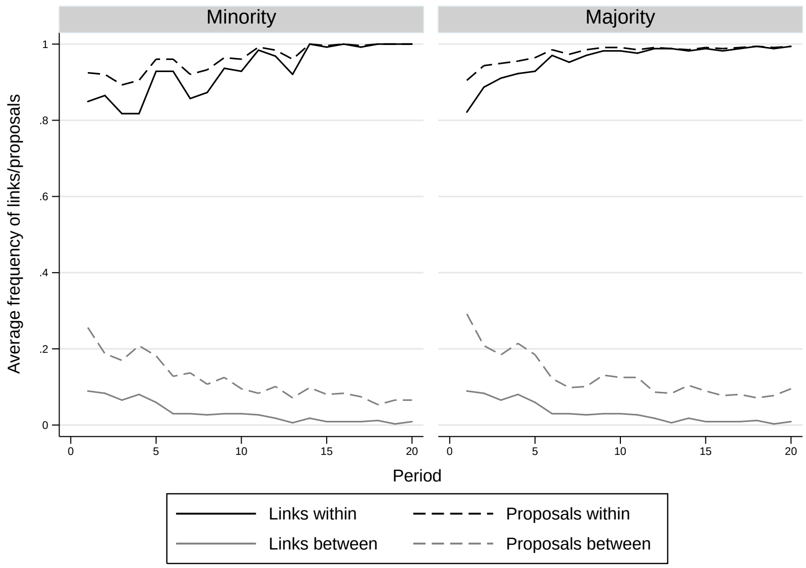


Figure 10: Link Choices in **ENDO**.

of actions. With an exogenous complete network, they converge gradually to conformism on the majority's preferred action.

There is a risk in forming links with individuals of the other type: it is not clear if successful coordination will be achieved and in the event of failure to coordinate, there is a negative return (equal to the cost of the link). To investigate the role of this risk in shaping behavior, we next turn to small and zero costs of linking.

3.4 The role of linking costs

We start with a discussion of a treatment with low linking cost, given by $k = 0.5$. The game is as in **ENDO**, but to make the payoffs comparable, we set the value of other parameters at $\alpha = 4.5$, $\beta = 2.5$. We refer to this low cost treatment as **CHEAP**.

To get a first impression of the pattern of linking and action choice Figure 11 presents the dynamics in one group. This figure suggests that lowering the costs of linking leads to greater linking across preference types, but that it does not significantly effect the long run outcome, relative to the baseline (High Cost) treatment: we still observe (almost complete) segregation and diversity.

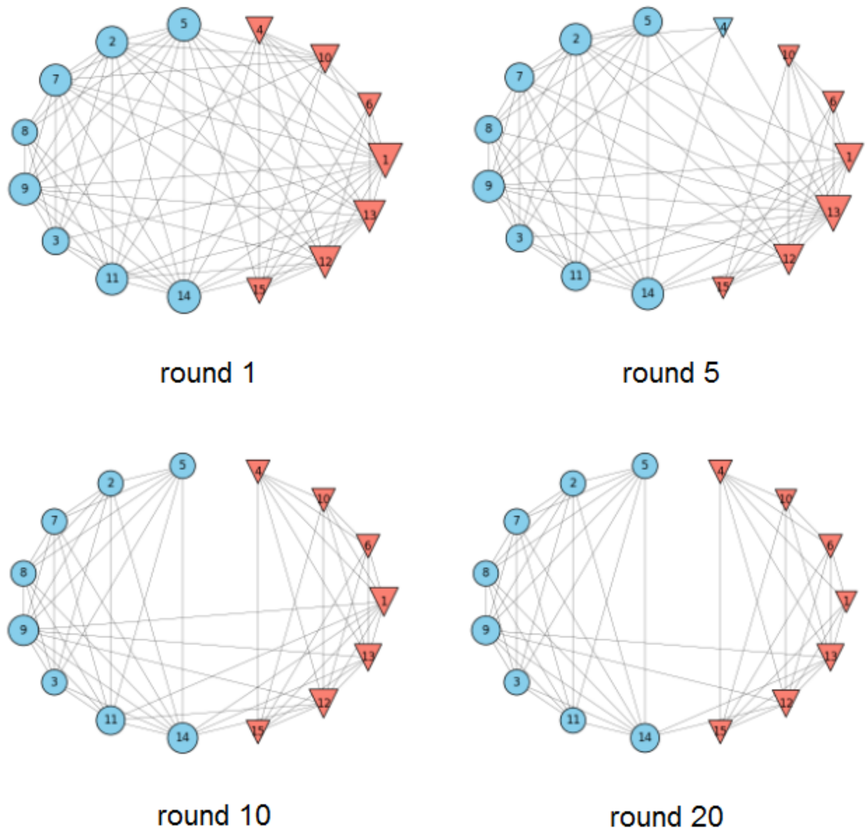


Figure 11: Illustration of dynamic behavior in **CHEAP**

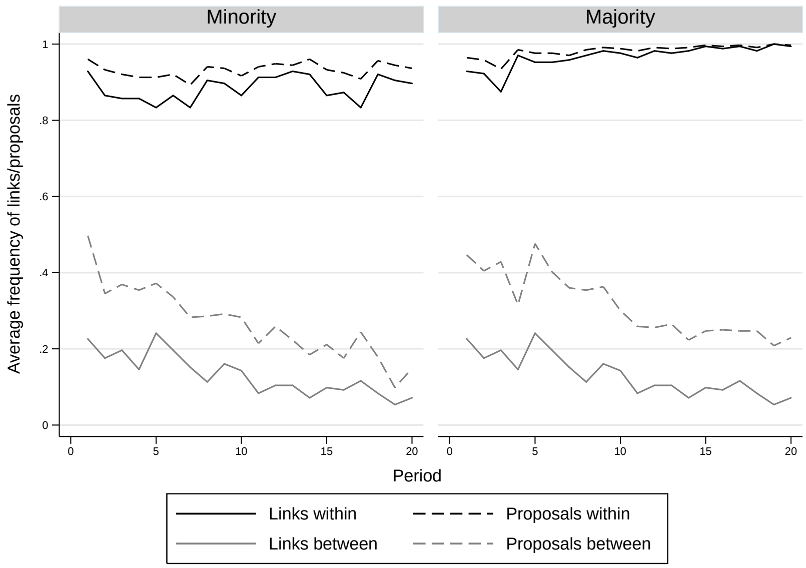


Figure 12: Link choices in **CHEAP**.

We turn next to the analysis of the data from the experiments. First, we discuss networks formed by subjects under **CHEAP** and we compare them to networks formed under **ENDO**. We observe that subjects in **CHEAP** very early on form all links within own group (over 80%) and very few links across groups (less than 20%). The members of the majority converge to almost full within group connectivity by round 10. Members of the minority exhibit less connectivity within own group: the fraction of all ties within the group remans around 90% toward the end of the experiment. But by round 10, across group links have fallen to less than 10% of possible links. And this persists until the end. The average degree in the majority's component is 6.94 (99% of the potential links are formed) and in the minority's component the average degree is 5.3 (88.3% of the potential links are formed). The between-type average degree is 0.62.

Next, we turn to the choice of actions. Figure 13 gives us a first impression of the dynamics of conformism. Members of the majority converge to their preferred action by round 5, and persist with this choice until the end of the rounds. The minority too chooses its preferred action by and large, and this trend emerges very early on and persists. Overall, less than 10% of the minority members choose the majority's preferred action eventually.

Finally, we consider a treatment with zero costs of linking. We refer to this treatment

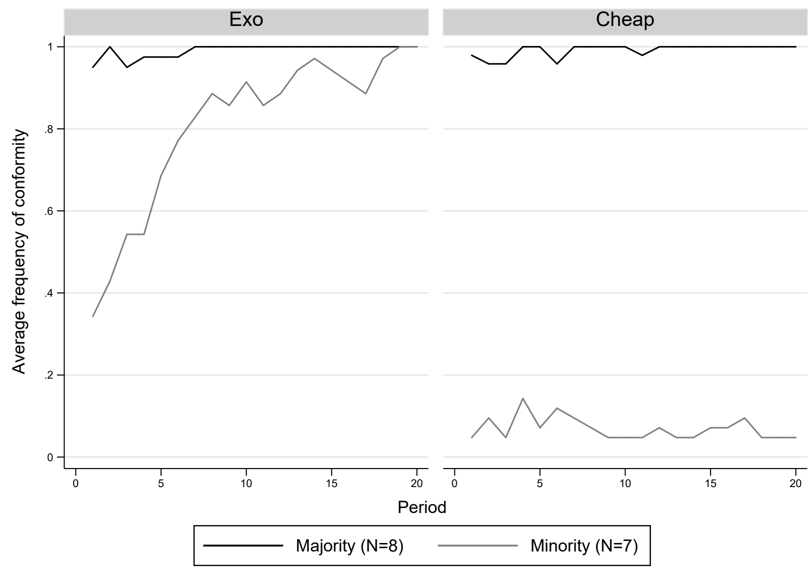


Figure 13: Action Choice in **CHEAP**.

as **FREE**. The values of the parameters are now set at $\alpha = 4$, $\beta = 2$ and $k = 0$. This mitigates the risk in linking fully: forming a link is now a 'weakly dominant' strategy. So we expect to see the complete network; this is the same network as in the exogenous complete network, and so we we conjecture that subjects would choose to conform on the majority action too, in line with the outcome under **EXO**.

To get a first impression of the pattern of linking and action choice Figure 14 presents the dynamics in one group. We notice that individuals are very active in creating links from early on: the network is close to being complete. Nevertheless, individuals choose their own preferred action: there is clear convergence to diversity in action. We now discuss the data from the experiments to get a more detailed picture of connectivity and action choice.

In the treatment **FREE**, having no linking costs promotes integration from the start (see Figure 15). The interaction structure is highly connected from round 1, and the high rates of connectivity continue over time, without much variation. Out of the 105 links that can be formed in all, across rounds there were on average 95.6 in **FREE** while there were only 49.13 in **ENDO** (z=-4.18, p=0.00). The majority formed more within-type links in **FREE** compared to **ENDO** (z=-3.31, p=0.0009), and so did the minority (z= -3.31,

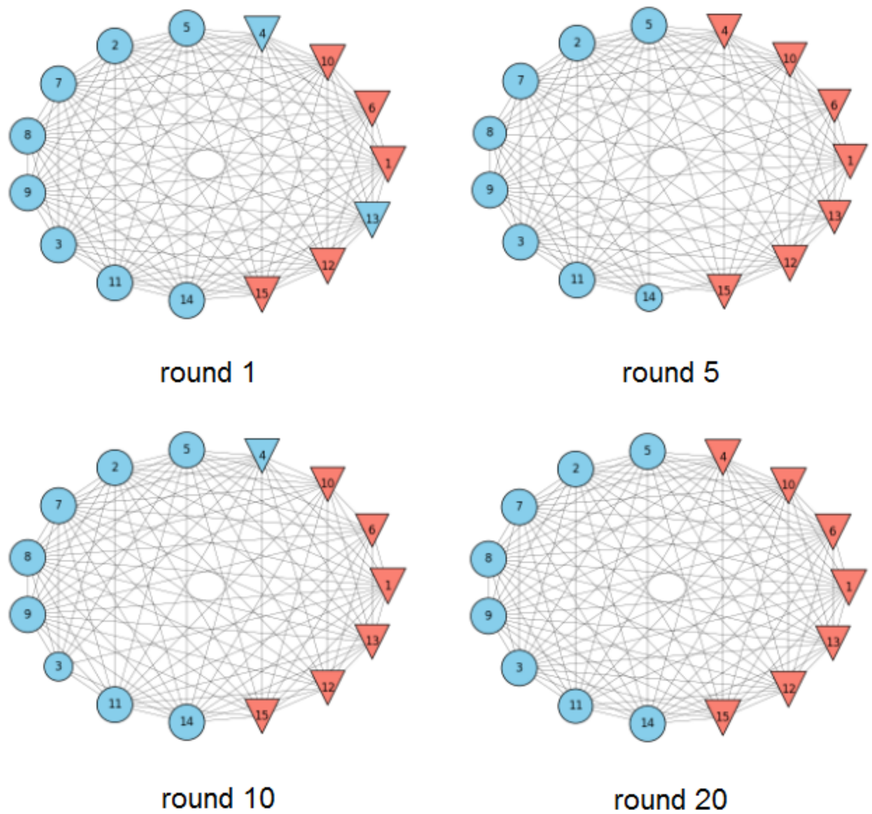


Figure 14: Illustration of dynamic behavior in **FREE**

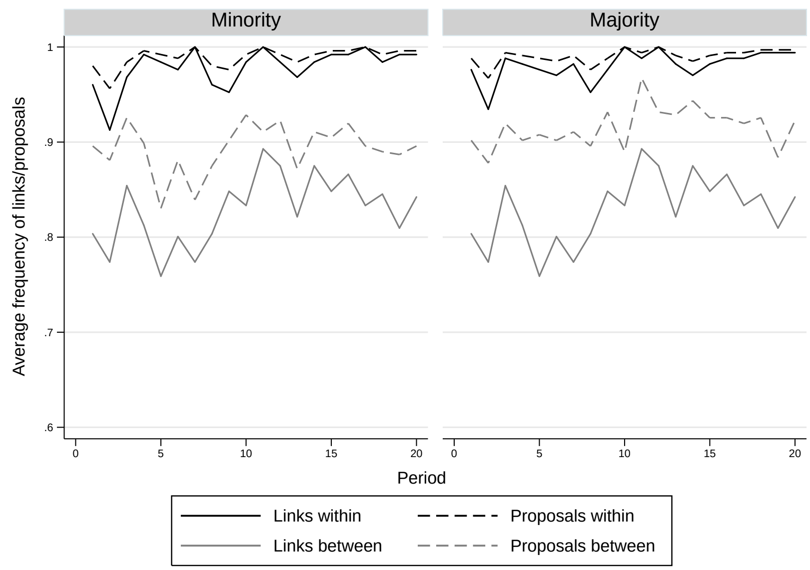


Figure 15: Link Choices in **FREE**.

p=0.0009). But the main difference was on across-group ties: this increased from 0.62 links in **ENDO** to 6.27 links in **FREE** (z=-4.8,p=0.00).

We then turn to action choice: we have seen that the network is close to being complete, so in line with the outcome under treatment **EXO** we should expect to see conformity on the majority's preferred action. Our main finding is that despite similarities in the interaction network, the behavior remains close to the segregated communities outcome. In other words, subjects choose distinct actions, even though they have formed links and are in a network that is close to a complete network. The majority chooses their most preferred action, as does the minority. Thus, we see *integration with diversity.*

Figure 16 illustrates the dynamics of action choice under **FREE** and compares it to the behavior observed under **ENDO**. The main observation is that the level of conformity is indistinguishable between **ENDO** and **FREE** (z=1.45, p=0.15). The coordination dynamics differ slightly: the majority converges faster to full coordination under **ENDO** as compared to **FREE** (z=4.42, p=0.00). But the diversity in behavior is clear cut, in spite of almost complete integration.

We note that the high level of connectivity in **FREE** leads to more attempts at coordination across types than in **ENDO** and **CHEAP**. The majority players choose the

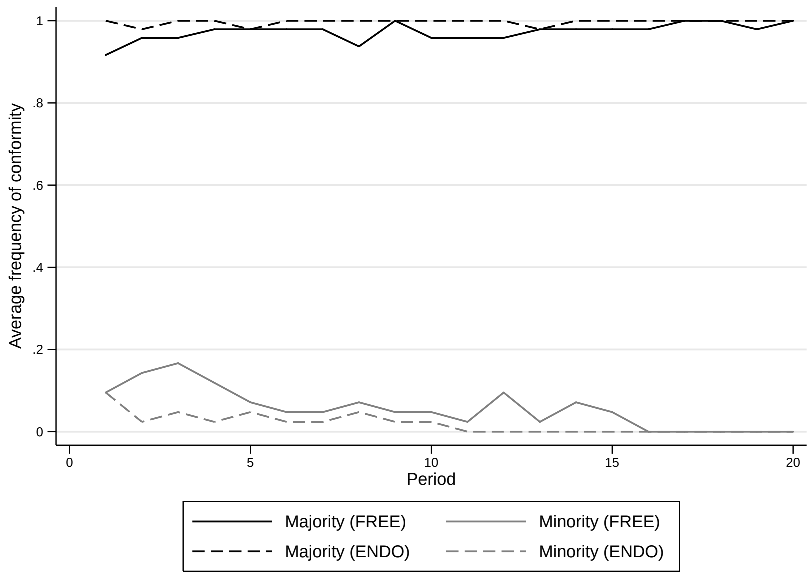


Figure 16: Action Choice in **FREE** vs. **ENDO**.

action preferred by the minority slightly more often, compared to what they would do under segregation. Similarly, the minority players tend to choose the majority action a little more often than in **ENDO**. But, in the aggregate, these attempts fail, and we observe, by round 15, a clear diversity of actions. To summarize:

Result 2. *When costs of linking are low but positive, subjects choose a segregated network and diversity of actions. When costs of linking are zero, subjects choose an almost fully integrated network, but persist with diverse action choices.*

To highlight the uniformly strong effects of endogenous linking we put together the behavior across the three linking cost treatments in Figure 17.

So far we have concentrated on the connectivity and the behavioral outcomes. Clearly, freedom of choice in linking has very strong and striking effects along these two dimensions. We now turn to the welfare implications. Table 2 provides an overview of the welfare effects. It shows the level of welfare attained in the laboratory as a percentage of the welfare attained in the first best outcome – conformism on the majority action.

We see that the overall loss is very large: society loses one-third of the efficient welfare. Interestingly, this loss is almost entirely borne by the majority, who lose almost one half of

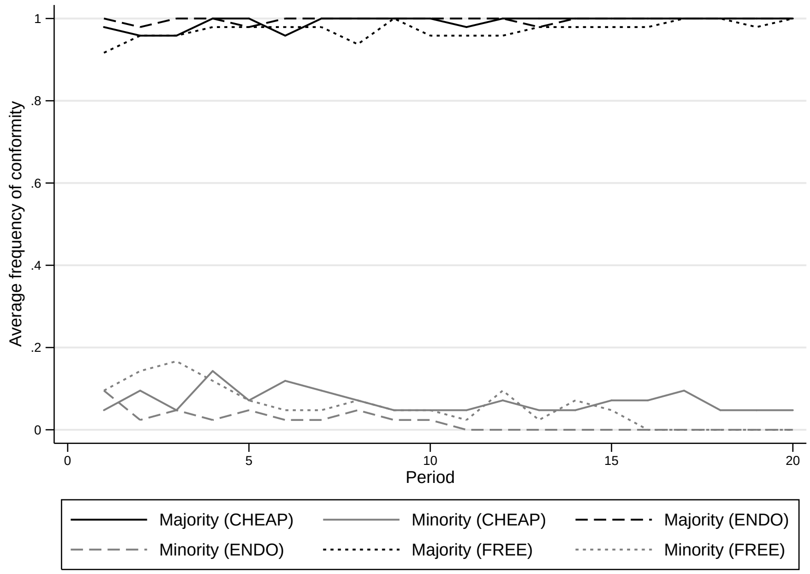


Figure 17: Action Choice in the endogenous linking treatments

Treatments (Outcomes)	All	Majority	Minority
EXO			
Conformity	100%	100%	100%
(Diversity)	(67%)	(55%)	(94%)
ENDO, **CHEAP**, **FREE**			
Segregation & diversity	67%	55%	94%
(Integration & conformity)	(100%)	(100%)	(100%)

Table 2: Freedom of association: effects on welfare

the attainable welfare (under integration and conformism). By contrast, the minority loses only 6% of their welfare (in the welfare maximizing outcome) by opting for segregation and diversity.[10]

[10]This difference in loss levels may suggest that inequality aversion is at work. However, observe that the equilibrium payoffs are exactly the same under exogenous as well as under endogenous networks. As we observe conformism and social optima in the exogenous setting, we rule out inequality aversion as an explanation for the segregation and diversity outcome.

3.5 Effects of minority size

We briefly study the effects of the size of the minority. As the size of the minority falls, the payoff losses of separating itself rise and we conjecture that this may induce greater integration and conformism.

To test this hypothesis, we conducted a final experiment in which we considered a majority of size 12 and a minority of size 3. We refer to this treatment as **LOWCONFLICT**. In this case, segregation remains a pairwise stable outcome, but the payoff losses for the minority in segregation outcome are very large. The game and parameters are as in **ENDO**; we only vary the group composition.

We find that the network is much more densely connected. Out of the 105-possible links that can be formed, individuals form on average 101.7 links across rounds, which is significantly greater than the level of connectivity in **FREE** (z=3.02, p=0.003). The majority formed significantly more links than the minority across rounds, 13.64 > 13.2 (z=-3.32, p=0.001). As illustrated in Figure 18, it took the minority longer to link within their own type than it took the majority. But more importantly, there were multiple unreciprocated proposals from the minority to the majority in the first round. On average, a member of the minority proposed 10.5 links to the majority in the first five rounds and only 7.6 links were formed. The majority, on the other hand, proposed 2.0 and formed 1.9 links with the minority in the same block. While the majority players were at first reluctant to create links with the minority, the minority players persisted with proposals. This persistence appears to have triggered reciprocity in the following rounds.

We observe that the initial reluctance of the majority to form connections with the minority breaks down due to both the persistence of link proposals from the minority and due to the behavior of the minority. More specifically, we observe that from the first round, 78% of the minority players conformed (i.e. chose their least preferred action), and by round 3 all minority players were conforming completely. The number of majority players choosing their preferred action is not different from 12 across rounds (t=-1.0, p=0.36) and the number of minority players conforming is exactly 3 in the last five rounds. Thus we conclude:

Result 3. *Consider the setting with* ***LOWCONFLICT****. With endogenous linking, subjects converge rapidly to complete integration and conformity on the majority's preferred action.*

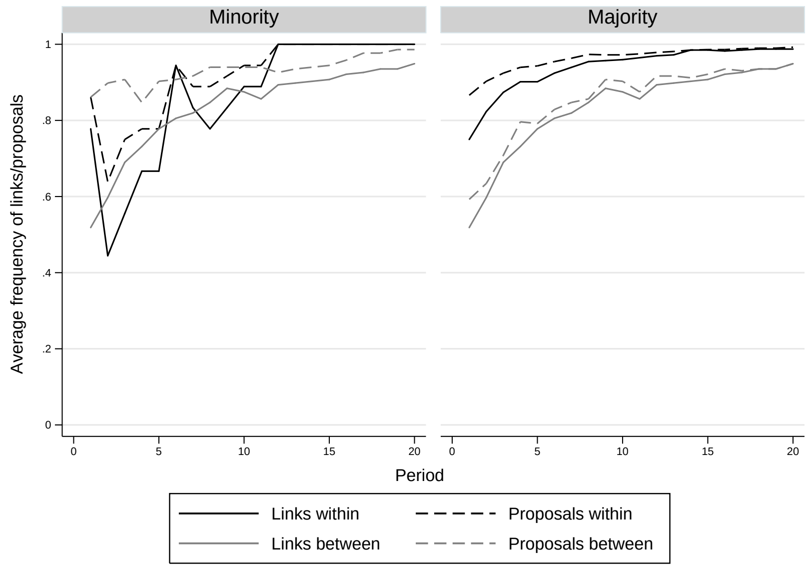


Figure 18: Link Choices under **LOWCONFLICT**

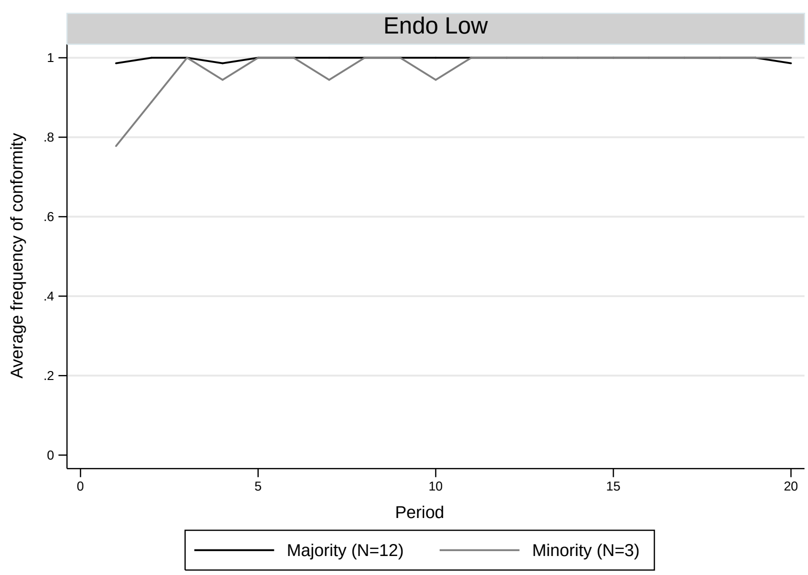


Figure 19: Action Choice under **LOWCONFLICT**

4 Theoretical Explanations

Our experimental findings point to the key role of linking activity by individuals in shaping their behavior. These findings motivate a closer theoretical examination of the process through which the linking matters. In this section we explore dominant approaches to understanding coordination problems that rely upon beliefs and dynamics and on introspection, respectively. We view these approaches as offering complementary explanations in helping us understand the experimental findings.[11]

4.1 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 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, $\mu^\epsilon(x)$

[11] 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 short 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 interaction, we feel that repeated game effects are not central to understanding this difference.

[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].

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 experiment, Proposition 6 specifies the existence of a unique stochastically stable state in the **EXO** treatment.

Proposition 6. *Consider an exogenous complete network. If $\frac{\beta}{\alpha} > \frac{n+4}{3n}$, then conformity on the majority's preferred action is the only stochastically stable outcome.*

The proof of this result is given in the Appendix. According to Proposition 6, 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^{t-1}_{-ij} for both i and j (such adjustment may therefore involve adding or severing the link $\overline{g}^t_{ij}$ and/or changing one or both actions x^t_i and x^t_j). Note that $\overline{g}^t_{ij} = 1$ implies that $x^t_i = x^t_j$ even if $x^{t-1}_i \neq x^{t-1}_j$. Similarly, $\overline{g}^t_{ij} = 0$ implies that $x^t_i \neq x^t_j$ even if $x^{t-1}_i = x^{t-1}_j$. (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 7. *Consider the endogenous linking model. If $k > 0$ and $k \geq 2\beta - \alpha$, then all pairwise stable equilibria are stochastically stable. If $k = 0$ and $\frac{\beta}{\alpha} > \frac{n+4}{3n}$, then integration with conformity on the majority's preferred action is the only stochastically stable outcome.*

The proof of this result is given in the Appendix. According to the first part of Proposition 7, stochastic stability under the assumption of the dynamic process described above remains indecisive and cannot predict any specific outcome whenever $k > 0$. More importantly, it is invariant to the size of the subgroups (N_a and N_b) and the linking cost $k > 0$, which have shown to significantly influence behavior in our experiment. Another noteworthy observation from the proof of the first part of Proposition 7 (in Appendix) is that the result is mainly driven by the possibility of trembles on the players' actions (no tremble on the link itself is necessary to reach the stability result). This therefore suggests that such a result is robust as it would hold despite assuming a slightly different dynamics as long as trembles are allowed on the actions.

However, the second part of Proposition 7 states that stochastic stability provides a clear prediction of convergence to integration with conformity on the majority's preferred action when $k = 0$. This is inconsistent with behavior observed in the **FREE** treatment.

To summarize: in an exogenous complete network, small perturbations in the dynamics lead to conformism on the majority's preferred action. This is consistent with the experimental finding. In the endogenous network model with positive linking cost, small perturbations sustain both integration and conformism, and a variety of segregation and diversity outcomes. The theoretical prediction is permissive: the experimental findings on segregation are weakly consistent with the prediction. However, in the absence of any linking cost, small perturbations lead to integration with conformity on the majority's action, which is not consistent with our experimental findings.

4.2 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 to accurately anticipate every other individual's behavior appeals for some salient mechanism to 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, according to which 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. In other words, 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] This assumption of same-type similarity in behavior considerably simplifies the decision problem in a fixed complete network, as shown through Proposition 8.

Proposition 8. *Suppose an exogenous complete network. If $|N_a| > |N_b|$ and $\frac{|N_b|}{n} < \frac{\beta}{\alpha} < \frac{|N_a|}{n}$, then conformity on the majority's preferred action is the only TR equilibrium.*

The proof of this result is given in the Appendix. Our empirical observations from **EXO** are consistent with Proposition 8. In particular, the above equilibrium is justified as follows: it is strictly dominant for the majority to play their preferred action, and knowing this, the minority is better off conforming to the majority's preferred action (see the proof of Proposition 8 for details). This difference in the depth of reasoning required highlights the difficulty for the minority to reach equilibrium as compared to the majority. Moreover, the same theory can be applied to the endogenous network formation game, as shown in Proposition 9.

Proposition 9. *If $|N_a| > |N_b|$ and $\frac{|N_b|}{|N_a|} < \min(\frac{\alpha-\beta}{\beta-k}, \frac{\beta-k}{\alpha-\beta})$, then:*

- *Integration with conformity on the majority's preferred action is the only TR equilibrium consistent with forward induction.*
- *Segregation with diversity is a TR equilibrium consistent with backward induction.*

[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 proof of this result is given in the Appendix. Let us assume that $r = \frac{\alpha-\beta}{\beta-k} = \frac{\beta-k}{\alpha-\beta}$, consistently with all our experimental treatments involving endogenous linking. In such scenarios, the socially efficient outcome is the unique TR solution of the game, which relies on some forward induction reasoning: the majority signals its intention to play its preferred action a by forming links with the minority. However, the salience of such a signal can affect the emergence of this solution. For example, if $\frac{|N_b|}{|N_a|} << r$, then the signal is very salient, and since the marginal benefit of the minority for being linked with the majority is also large, integration with conformity is likely to emerge. This is consistent with our observations from **LOWCONFLICT**. On the other hand, as $\frac{|N_b|}{|N_a|}$ becomes closer to r,[15] the signal from the majority weakens, and the minority becomes more indifferent about whether to form links with the majority. As a result, convergence to segregation with diversity is more likely to emerge as it also characterises a (risk-free) sub-game perfect TR equilibrium. This is consistent with our observations in **ENDO** and **CHEAP**. In the particular case where $k = 0$ as in **FREE**, although Proposition 9 still applies, there exist alternative TR equilibria where linking occurs across different types and everyone selects their preferred action. In fact, suppose that every minority player forms links with 7 out of the 8 majority players (resulting in a total of 49 links out of 56 possible links across types). In this case, conforming to the majority's preferred action would yield each minority player a payoff of 28, which is also the payoff for selecting their preferred action. As a result, diversity in such a highly connected network configuration can hold as a possible TR equilibrium. [16] Our observations in **FREE** are therefore consistent with this kind of equilibrium behavior (according to Figure 16, less than 90% of links across types are formed).

To summarize: the theory of team reasoning sketched above points to two features that facilitate integration and conformity with endogenous linking: one, a sufficiently strong signal of the majority's intention to play their preferred action through linking, and two, a sufficiently large marginal benefit for the minority to form links and conform.

4.3 Social preferences and bounded reasoning

In this section we briefly discuss two other possible theories: social preferences and level-k reasoning.

[15] Note that if $\frac{|N_b|}{|N_a|} > r$, then the forward induction argument does not hold anymore (see details in the proof of Proposition 9 in the appendix).

[16] Note however that this kind of TR equilibrium does not hold whenever $k > 0$.

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. This is motivated by the observation that conformism creates a large gap in payoffs between the minority and the majority, whereas payoffs are relatively similar under heterogeneity. We note that this argument applies equally well for exogenous and for endogenous treatment. 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 segregation and 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.

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.

5 Conclusion

The aim of this paper was to study the relation between two values: freedom of association and social cohesion.

We start by defining a network formation and action choice game in which individuals benefit from selecting the same action as their neighbours. However, one group of individuals prefers to coordinate on one action, while the other group prefers to coordinate on a second action. There exist multiple equilibria, which come broadly in two forms: one, integration where the network is fully connected, and every player conforms to the same action and two, segregation where the network is composed of two groups, with members of the groups choosing a distinct action. We show that social welfare is (uniquely) maximized under full integration and conformity on the majority's action.

Faced with the multiplicity in equilibrium, we conduct laboratory experiments with human subjects. We observe that individuals with different preferences segregate into distinct groups with different actions. This diversity of actions is robust with respect to the costs of linking.

To understand the role of freedom of association, we then examine a setting with an exogenous network. Again in the theoretical model, there exist a variety of equilibria, displaying conformism and diversity. But, in the experiment we observe that subjects almost always choose to conform on the action preferred by the majority. We therefore conclude that there is a tension between two values: social cohesion and freedom of association. This tension can lead to large welfare losses.

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

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. An "evolutionary" interpretation of van huyck, battalio, and beil's experimental results on coordination. Games and Economic behavior, 3(1):25–59, 1991.

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

J. A. Cuesta, C. Gracia-Lázaro, A. Ferrer, Y. Moreno, and A. Sánchez. Reputation drives cooperative behaviour and network formation in human groups. Scientific reports, 5: 7843, 2015.

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.

E. Gallo and C. Yan. The effects of reputational and social knowledge on cooperation. Proceedings of the National Academy of Sciences, 112(12):3647–3652, 2015.

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

P. Hernandez, M. Muñoz-Herrera, and A. Sanchez. Heterogeneous network games: Conflicting preferences. Games and Economic Behavior, 79:56–66, 2013.

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.

M. Kosfeld, A. Okada, and A. Riedl. Institution formation in public goods games. The American Economic Review, 99(4):1335–1355, 2009.

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. 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.

J. Wang, S. Suri, and D. J. Watts. Cooperation and assortativity with dynamic partner updating. Proceedings of the National Academy of Sciences, 109(36):14363–14368, 2012.

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

Appendix A: Proofs

PROOF OF PROPOSITION 6:

Let N_{maj} be the majority group whose members prefer action $x \in \{a, b\}$, i.e., $N_{maj} = \{i \in N : \theta_i = x\}$, and $N_{min} = N \backslash N_{maj}$ represents the minority group in N ($|N_{min}| = 7$). 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 ($\in \{a, b\}$), 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]. $\square$

PROOF OF PROPOSITION 7:

Suppose that $k > 0$. 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, it follows from Proposition 1 that a player in any connected group M chooses his preferred action if there are at least m other players from M to choose the same action, where $m > \frac{\beta}{\alpha+\beta}|M| - \frac{\alpha-\beta}{\alpha+\beta}$. It is then easy to see that whenever $M = \{i\}$, the isolated player i must be playing his preferred action ($x_i^t = \theta_i$), and this outcome is stable as long as $2\beta - k \geq \alpha$. Let us now assume that $|M| > 1$ and $x_i^t \neq \theta_i$ for some $i \in M$. In this case, i can benefit by deleting a link within M and switching to $x_i^t = \theta_i$ only if $|M|\beta - k < \alpha$. Alternatively, i can benefit by adding a link with a player $j \notin M$ and switch to $x_i^t = x_j^t = \theta_i$ if $|M| < \frac{2\alpha-k}{\beta}$. If $|M|\beta - \alpha < k$ or $|M| < \frac{2\alpha-k}{\beta}$, the corresponding state is stable only if all players in M are choosing their preferred actions, i.e., $M \subseteq \{i \in N : x_i^t = \theta_i\}$. However, whenever $|M|\beta - \alpha \geq k$ and $|M| \geq \frac{2\alpha-k}{\beta}$, then any possible configuration is stable (players' types within each component are irrelevant).

Let us now determine the recurrent communication classes. We denote C_{maj} as the integration state with conformity on the majority's action, and C_{min} as the integration state with conformity on the minority's action. Moreover, it is clear from above that there exist multiple segregation states. However, looking at the resistance between pairs of segregation states, we observe that, for any segregation state seg_1, there exists another segregation state seg_2 such that $r(seg_1, seg_2) = 1$. In fact, a single tremble on an action in seg_1 suffices to reach seg_2's basin of attraction. More specifically, take a segregation absorbing state involving two complete components M_1 and M_2. In this case, if a tremble leads a player $i \in M_1$ to switch action, then all other members from $M_1 - \{i\}$ will be better off deleting their link with i. Moreover, since adding links between i and any member of M_2 would be mutually beneficial, links will be formed between i and all members of M_2. This results in a new segregation state involving two complete components $M_1' = M_1 \backslash \{i\}$ and $M_2' = M_2 \cup \{i\}$. Moreover, a similar argument implies that $r(C_{maj}, seg_1) = r(seg_1, C_{maj}) = r(C_{min}, seg_2) = r(seg_2, C_{min}) = 1$ for some segregation states seg_1 and seg_2 involving an isolated player ($M_1 = \{i\}$ and $M_2 = N \backslash \{i\}$). Thus, we have that for any absorbing state s_1, there exists another absorbing state s_2 such that $r(s_1, s_2) = 1$. It follows that, for any absorbing state x, the corresponding x-tree is a directed graph where the resistance of every edge is exactly 1. As a result, the resistance of s is $r(x) = n - 1$. Since all absorbing states carry the same resistance, it directly implies that all absorbing states are stochastically stable.

Now suppose that $k = 0$. Again, integration with conformity on the same action is still stable, and therefore the corresponding recurrent communication classes C_{maj} and C_{min} still hold. It is then easy to show that the only other absorbing state consists in the complete network where every player selects their preferred action. In fact, since $k = 0$, it is weakly dominant for everyone to form links with everyone else, in which case such diversity is a stable outcome. We refer to this recurrent communication class as D. Note that any stable segregation outcome from above where some players play their least preferred action belongs to D. The proof of Proposition 6 then directly follows. □

PROOF OF PROPOSITION 8:

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 $|N_a|\alpha$ for playing a, and at most $n\beta$ for playing b. If $\frac{\beta}{\alpha} < \frac{|N_a|}{n}$, then it is strictly dominant for the majority to play a. Moreover, the minority would then obtain $|N_b|\alpha$ for selecting b, and $n\beta$ for selecting a (assuming the majority plays a). Since $\frac{\beta}{\alpha} > \frac{|N_b|}{n}$, the minority is then strictly better off selecting a. This yields conformity on the majority's action as the only equilibrium solution. □

PROOF OF PROPOSITION 9:

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. Regarding the forward induction solution, the reasoning is as follows. The majority would obtain $|N_a|(\alpha - k) + k$ for not forming links with the minority (and playing a). If the majority instead links with the minority, then playing b would yield $|N_a|(\beta - k) + k - |N_b|k$ if the minority plays a, or $n(\beta - k) + k$ if the minority plays b. Since $\frac{|N_b|}{|N_a|} < \frac{\alpha-\beta}{\beta-k}$, forming links with the minority and playing b is strictly dominated by not forming links and playing a. As a result, if the majority forms links with the minority, it will then play a (forward induction). Similarly, the minority would obtain $|N_b|(\alpha - k) + k$ if not forming links with the majority (and playing b). However, if the minority links with the majority, then playing a would yield $n(\beta - k) + k$, and playing b would yield $|N_b|(\alpha - k) + k - |N_a|k$ (the majority would then play a as shown above). Since $\frac{|N_b|}{|N_a|} < \frac{\beta}{\alpha-\beta}$, if the minority forms links with the majority, it would be strictly better off selecting a. Moreover, since $\frac{|N_b|}{|N_a|} < \frac{\beta-k}{\alpha-\beta}$, the minority is strictly better off forming links with the majority and play a. This leads to integration with conformity on the majority's preferred action. $\square$