Discussion Papers in Economic Behaviour

DPEB
01/14

Conflict and segregation in networks: An experiment on the interplay between individual preferences and social influence

>Lea Ellwardt
ICS, Faculty of Behavioral Social Sciences, University of Groningen, The Netherlands

>Penelope Hernández
Universitat de València and ERI-CES, Spain

>Guillem Martínez-Cànovas
Universitat de València and ERI-CES, Spain

>Manuel Muñoz-Herrera
ICS, Faculty of Behavioral Social Sciences, University of Groningen, The Netherlands

www.erices.es

Conflict and segregation in networks: An experiment on the interplay between individual preferences and social influence

By Lea Ellwardt, Penélope Hernández, Guillem Martínez-Cánovas and Manuel Muñoz-Herrera*

We examine the interplay between a person's individual preference and the social influence others exert. We provide a model of network relationships with conflicting preferences, where individuals are better off coordinating with those around them, but not all prefer the same action. We test our model in an experiment, varying the level of conflicting preferences between individuals. Our findings suggest that preferences are more salient than social influence, under conflicting preferences: subjects relate mainly with others who prefer the same. This leads to two undesirable outcomes: network segregation and social inefficiency. The same force that helps people individually hurts society.
JEL: C62, C72, D82, D85
Keywords: Heterogeneity, Social Networks, Formation, Equilibrium selection

The interplay between what we prefer to choose and the influence those around us exert on our choices is at the core of our social and economic life. Both individual preferences and social influence guide our behavior and whether to establish relationships with others or not (Tajfel and Turner 1979, Lazarsfeld, Merton et al. 1954, McPherson, Smith-Lovin and Cook 2001). For instance, when choosing our friends (Marsden 1990) or neighbors (Schelling 1978) individual preferences are a strong determinant of how we make such decisions. But also, the social influence peers exercise on human behavior is enormous (Jackson 2009), affecting whether people act in alignment or not with those they relate to (Morris 2000, López-Pintado 2006). Examples of social influence range from which products we buy or languages we learn (Galeotti et al. 2010), whether we engage or not in criminal activities (Ballester, Calvó-Armengol and Zenou 2006), to our participation in collective action (Granovetter 1978). Arguably, by addressing the interplay between individual preferences and social influence we can understand the forces motivating how people decide what rela-

* Ellwardt: ICS, Faculty of Behavioral Social Sciences, University of Groningen, Grote Rozenstraat 31, 9712 TG, Groningen, The Netherlands, l.ellwardt@rug.nl. Hernández: ERI-CES, Departamento de Análisis Económico, Universitat de Valencia, Avenida Los Naranjos s/n, Campus de Tarongers, 46022, Valencia, Spain, penelope.hernandez@uv.es. Martínez-Cánovas: ERI-CES, Departamento de Análisis Económico, Universitat de Valencia, Avenida Los Naranjos s/n, Campus de Tarongers, 46022, Valencia, Spain, guillem.martinez@uv.es. Muñoz-Herrera: ICS, Faculty of Behavioral Social Sciences, University of Groningen, Grote Rozenstraat 31, 9712 TG, Groningen, The Netherlands, e.m.munoz.herrera@rug.nl.

tionships to form and how to behave with others, that is the aim of this paper.

One of the most prominent theoretical tools to study the effect individual preferences have on the way people behave is *identity theory* (Tajfel and Turner 1979, Akerlof and Kranton 2000). From the perspective of identity theory a person's sense of self, her identity, is composed by three elements. First, *categorization*, putting ourselves and others into social categories (i.e., being a Christian orthodox, a female, a police man). Second, *identification*, the process we use to associate ourselves with certain groups. The group we identify with, say because we share a common identity with its members, is the *in-group*. Conversely, the group we do not identify with, for we do not share the identity of its members, is the *out-group*. Third, *comparison*, the process we use to compare our in-group and the out-group, most likely favoring one over the other. Identity theory has highlighted how the social categories people identify with are associated with particular behaviors prescribed for them. We refer to this prescribed behavior as a person's individual preference. Thus, what people care about and how much they care about it greatly depends on their identity. For example, in latin cultures, when dancing salsa or tango, males are meant to lead and females are supposed to follow; such is the behavior associated to each category. In this direction, identity theory stresses that a person obtains greater benefit from behaving as indicated by her identity than doing otherwise. When people are doing what is in accordance to their individual preferences they are happy, they get more out of it, and those who are not living up to the norms set by their social categories are unhappy, so they tend to change their decisions to meet their standards (Akerlof and Kranton 2010).

On the other hand, a leading research program studying how the structure of social relationships influences behavior is that of *strategic interaction in networks* (i.e., network games). Work on network interactions gives account of the way we make our decisions influenced by the decisions of our neighbors. For instance, if a person is choosing a technological product and wants it to be compatible with her co-workers or friends, her choice can change depending on how many of them are using the same technology or a different one (Vives 1990, Vives 2005). These interactions are known as coordination games with strategic complementarities, where a person's incentives to choose a given product or adopt a given behavior increase as more of those around her make the same choice. The underlying mechanism from social influence is that people perceive coordinating with the behavior of others as beneficial for them. As a result, this line of research has highlighted that people are more likely to adopt a given behavior or not depending on who they are related with, even if such a behavior is not the one prescribed for their identity (Hernández, Muñoz-Herrera and Sánchez 2013).

The existing research on these two lines of work has illustrated ways in which identities *or* social influence affect our relationships and our behavior. However, it leaves open the very fundamental aspect of how these elements relate to each other and work together. The current paper aims to address this gap and give account of the interplay between individual preferences and social relationships. To do so, we elaborate and analyze a formal model where actors choose with whom to interact and which behavior to adopt (i.e, network games), and experimentally tests the model by varying the way identities and social influence take place. Our model moves beyond the existing work in its combination of three features. First, our model introduces identities as part of the strategic considerations actors have by allowing for heterogeneity in social categories. In our case there are two social categories and an actor either belongs to one or the other. Second, to assess the effect of identities on the establishment of relationships, actors in our model form a social network by making decisions about whom to link with and whom to leave out. Particularly, the choice of forming connections is made after actors are informed of their own identity and the identity of the other actors in the population. Third, to understand how social influence affects actors choices, we model the adoption of behavior as a choice that is made once the structure of relationships has been formed. There is one behavior prescribed to each social category, so that the preference of an individual is to adopt the behavior that corresponds to her identity but there is a benefit in behaving the way those around us do. In this way, our theoretical model considers the essentials of identity theory and social influence in network relationships to unravel the way these two determinants of our decision-making process relate to each other.

A key aspect of the relationships we model is that they portray strategic complementarities. This means that actors are better off aligning their behavior to that of those around them (i.e., their network relationships). But, by introducing identities actors are in conflict about the behavior each prefers to adopt. Depending on the social category they belong to, some actors prefer one behavior and others prefer a different, yet they rather coordinate with as many others as possible. Thus, we model the interplay between identities and social influence in a context of *conflicting preferences.* In this setting we design an experiment in which subjects choose with whom to connect and how to behave playing a game derived from our theoretic model. In our experimental design subjects are artificially assigned an identity and they know the identities of others. Our focus is to consider different conditions where the relative size of the social categories vary. In this way we are able to control the social context and therefore the intensity of the conflict in preferences between social categories. On one hand, social influence points to the idea that subjects rather convey to the pressure of the strongest category (i.e., the majority) and be better off by it. On the other, identities showed that subjects will require different levels of pressure to adopt the behavior that does not correspond to their individual preference, for people have a strong inclination to behave accordingly to the prescription for their social category. In this way, our theoretical and experimental work contribute to the understanding of how the interplay between identities and social influence determine what relationships are formed and what behaviors are adopted in a network environment.

The remainder of this paper builds as follows: In section I we describe the theoretical framework of our modeling and experimental design in relation to previous research on identity theory and on social influence. Our game theoretic model is presented in Section II. Section III analyzes the network structures that emerge from the interactions of actors belonging to different social categories, and the conditions under which either identities or social influence are stronger determinants of behavior and of the resulting network architectures. In section IV we describe the experimental study, the design, procedures and methods used. Section V presents the main results of our study guided by hypotheses derived from our theoretical model. We conclude with a discussion of the implications and limitations of the study in Section VI.

I. Theoretical framework: Identities and social influence

Our theoretical framework builds on two lines of work examining how relationships and behavior emerge from actors' individual preferences and the influence from those around them: *identity theory* (originating from psychology but recently increasingly adopted in economics; see (Akerlof and Kranton 2010) and *strategic interaction in networks* (from economics). While elaborations of these lines of work differ in the extent to which actors are modeled as perfectly or imperfectly rational and strategic (i.e., myopic or farsighted), both identity theory and the theory of strategic interaction in networks start from the assumption that individual actors strive to obtain optimal outcomes for themselves. Our study integrates both lines of research for the particular case of interactions with *strategic complementarities*.

Research on the theory of identities was initiated in psychology (Abrams and Hogg 2012, Tajfel 1978, Tajfel and Turner 1979, Turner et al. 1987), mainly focusing on the effects that the social context has on group processes and intergroup relations. The aim, to understand how different inter-group interactions could be explained and whether groups of people who share/differ in certain traits were more likely to integrate or discriminate each other (Tajfel and Turner 1979). A consistent finding in identity theory is that people favor their in-group relative to out-groups, because people desire a positive and secure self-concept, which leads them to think of their groups as good groups. The argument of in-group bias has been widely supported by experimental research on identities (Billig and Tajfel 1973).

To assess the effect of identities on inter-group relations, experimental studies on identities are characterized for their use of a methodology called the *minimal group paradigm.* In these experiments researchers sought minimal conditions that would create group identification. To do so, subjects were assigned to groups using arbitrary criteria (i.e., the toss of a coin). After informing subjects of their group membership (i.e., their identity), they were asked to allocate points to members of their own group (the in-group) and to members of the other group (the out-group). Minimal group experiments have typically shown a tendency to allocate more points to in-group members than to out-group members (Brewer 1979, Mullen, Brown and Smith 1992). This tendency of maximum differentiation between in-group and out-group has even occurred when it means sacrificing absolute in-group benefit. Psychological research on identities has illustrated the strong tendencies that group identification generate on our individual preferences. Nonetheless, this findings do not come without shortcomings. An important limitation is that this approach has no strategic considerations about the way people behave given the behavior of others. In general, participants in these experiments could not benefit or lose in any way from their point allocation strategy, and even in some experiments points did not carry any value at all (Turner 1978). Therefore, the importance of identities for the understanding of rational behavior was not clear from the existing research in psychology. The interaction of identity considerations and individual incentives had not been directly addressed theoretically or experimentally, leaving an important gap for the development of rational choice theory.

Goerge Akerlof and Rachel Kranton initiated research on identities in economics by developing a model in which identities are introduced in the utility function of the actors (Akerlof and Kranton 2000). By doing this, they were able to characterize ways in which identities are included as part of the process of maximization when rational actors choose how to behave. For instance, an action may increase monetary benefits but decrease identity utility, such as complying with social pressure to behave as opposed to the prescription of our social category. The application of their model has been found useful to explain gender discrimination (Akerlof and Kranton 2000), education (Akerlof and Kranton 2002), and contract theory (Akerlof and Kranton 2005). A set of experimental work has also included identity as part of the analysis, addressing the limitation that the psychological approach has in the perspective of the behavior of a rational actor, by taking into account monetary stakes (Bernhard, Fehr and Fischbacher 2006, Goette, Huffman and Meier 2006, Tanaka and Camerer 2009, Eckel and Grossman 2005, Charness, Rigotti and Rustichini 2007, McLeish and Oxoby 2007, Chen and Li 2009). Particularly, (Chen and Li 2009) have adopted the minimal group paradigm and showed that group divisions matter even when monetary stakes are involved. Subjects gave more points to members of their in-group, and in cases where punishment was possible they punished out-group members more. While the existing modeling of identities in economics provides insight into broad patterns of social behavior, it does not incorporate the micro-details of who interacts with whom (i.e., social networks). The inclusion of network relations in the analysis is a matter of great importance because networks have a profound effect on our decision-making process, and have proven to be necessary for our understanding of the way others influence our behavior.

Research on network interactions has introduced the strategic behavior of people into the analysis of social influence by modeling the interaction as a game (for surveys of the literature see (Goyal 2007, Jackson 2009, Vega-Redondo 2007). Network games model the way individuals behave as a function of the actions of their neighbors. For the case of coordination games with strategic complementarities; settings where individuals are better off the more of their neighbors in the network behave as they do but there are at least two possible behaviors, network research has captured individual behavior through thresholds (Granovetter 1978, Galeotti et al. 2010). Such thresholds are the representation of the social influence a person requires from her neighbors to adopt a given behavior. For instance, when a person is deciding whether to acquire a specific technology or not, if more than a given number of her neighbors (i.e., the threshold) have that same technology, this person would acquire it as well, otherwise she would acquire a different one. A main interest in this line of research has been to understand equilibrium selection, for there are multiple equilibria and it is not clear which outcome is more likely to occur. It is possible that all actors choose one of the available options, the same for all, or some acquire one technology and some acquire the other. Work following this aim are (Ellison 1993), (Kandori, Mailath and Rob 1993), (Young 1993), (Morris 2000), and (López-Pintado 2006). A persistent finding in the theoretical modeling of social influence in games with strategic complementarities is that the most likely outcome is the risk-dominant equilibrium. This means that instead of aiming to get the highest payoffs by choosing a risky option, actors are more likely to focus on the less risky behavior at the expense of payoffs. Two main aspects of this research that need attention are: (i) relationships are given exogenously, so that people do not have the choice of selecting with whom they want to interact, and (ii) actors have been assumed to be identical so that identities are not part of the analysis.

The first aspect of these limitations has received a great deal attention by modeling social relationships as endogenous decisions actors make (Jackson and Wolinsky 1996). This block of research aims to understand which network structures will emerge when rational actors have the discretion to create and severe their connections. Papers following this aim are (Jackson and Wolinsky 1996), (Bala and Goyal 2000), (Jackson and Watts 2002) and (Muñoz-Herrera et al. 2013). A main finding that endogenous formation brings to network games is that the risk-dominant equilibrium is not the most salient equilibrium anymore. So that if actors can choose with whom they want to affiliate, other outcomes are likely. The possibility actors have to select their partners reduces risk and the payoff dominant equilibrium becomes salient (Jackson and Watts 2002). Social influence has therefore a strong impact on behavior given the network of relationships. But, the possibility people have to influence the relationships they form proves to have a strong impact on the outcomes that occur. The idea is that people act strategically when deciding with whom to form social relationships. We choose our relationships because they are beneficial to us, and if an existing relationship with someone is not beneficial anymore, it is very likely to terminate it (Jackson and Wolinsky 1996). Thus, a particularity of social influence is that its strength can vary depending on whether we are able to adapt our behavior to respond to what others around us are doing or to adapt our relationships with others given what we are interested in choosing.

The second aspect of these limitations, the inclusion of identities, has not received much attention until now. A study of conflicting preferences, closely linked to ours, is the work by (Hernández, Muñoz-Herrera and Sánchez 2013). In their model the authors address the effect of heterogeneity in identities in network games. However, their analysis is restricted to a particular set of exogenously given networks (i.e., Erdös-Renyi networks), so that actors have no choice regarding whom they relate to. Our model extends (Hernández, Muñoz-Herrera and Sánchez 2013) into a two stage game in which actors endogenously decide over their connections in the first stage and then play a coordination game with strategic complements in the second stage. Our extension is motivated by the pervasive empirical findings showing how actors' identities influence who they connect with in their networks. For instance, many social networks portray homophily (Jackson 2009) and show that is it more likely to have friends of the same race (Marsden 1990) or gender (Verbrugge 1977). By modeling both stages we can study how the level of conflicting preferences influences the way networks are strategically formed, given the interplay between individual preferences and social influence.

II. The model

In this section we present our model of network interactions taking into account identities and social influence. Identities are associated to two social categories, each giving a behavioral prescription for the players, and the utility from the adopted behavior will depend on the identity of the players. This means that in our network game players have identities, each identity is associated with a behavior that gives it higher payoffs than the other, and the identities and behavior need not be the same for all players. Thus, conflicting preferences can be present as part of the social interaction.

Consider the set of players $N = \{1, \ldots, n\}$, with cardinality $n \geq 2$, who interact in a network game denoted by Γ. In Γ there are two social categories expressed by the set $\Theta = \{0, 1\}$. Every player $i \in N$ is *ex-ante* and exogenously endowed with an identity corresponding to one of the two social categories, $\theta_i \in \{0, 1\}$. Prior to the start of the game, players are informed about the size of the network and the identity of all players, including theirs. The network game Γ has two stages: affiliation and behavior adoption.

In the first stage, *affiliation*, players decide with whom they want to interact in the game. To do so, players create undirected connections between them. These connections are only created if both players mutually agree on their formation. Therefore, the action set of player i is a vector in $\{0, 1\}^N$. We denote by $\mathbf{p^i}$ the vector of connections proposed by player i at stage 1 where $p^i_j = 1$ means that player i proposes a link to j, and $p^i_j = 0$ otherwise. We suppose that $p^i_i = 0$. Only if $p^i_j = p^j_i = 1$, we say there is a link between i and j. The profile of vectors

$\mathbf{p} = (\mathbf{p^1}, \mathbf{p^2}, \ldots, \mathbf{p^n})$ represents the network by the set of links, g. Notice that, the set of potential connections is the complete network, g^N, and any network configuration is part of the set $G = \{g : g \subset g^N\}$. In the network, if a pair of players i and j are connected by a link, it is denoted as $g_{ij} = g_{ji} = 1$, and if there is no link between them, we say $g_{ij} = 0$. The set of neighbors a player i has is $k_i(g) = \{j : g_{ij} = 1\}$, $\forall\ j \neq i$. For simplicity we assume that $ii \notin g$, so that all neighbors in $k_i(g)$ are different from i. The cardinality of $k_i(g)$ is k_i, the degree of node i in the network.

In the second stage of the game: *behavior adoption*, players choose an action from the binary set $X = \{0, 1\}$, once the network has been formed. The action chosen by i, $x_i \in X$, is the same for all neighbors she plays with. We construct identity-based preferences given the existing social categories. A player i who has identity 1 (0) *prefers* action 1 over 0 (0 over 1). This is a behavioral prescription expressed in the payoff function below. We denote $x_{k_i}(g)$ as the vector of actions taken by i's neighbors. The game is expressed through a linear payoff function, u_i, that strategically depends on the choices made by connected players (i.e., those that can influence i's behavior), their identities and proposed links in the first stage, as follows:

$$u_i(\theta_i, \mathbf{p}, x_i, x_{k_i}(g)) = \lambda_{x_i}^{\theta_i} \left(1 + \sum_{j=1}^{k_i} I_{\{x_j = x_i\}} \right) - c \sum_{j=1}^{n} p_j^i, \tag{1}$$

where $I_{\{x_j = x_i\}}$ is the indicator function of those neighbors choosing the same action as player i. The parameter λ is defined by $\lambda_{x_i}^{\theta_i} = \alpha$ when a player chooses what she likes ($x_i = \theta_i$), the action prescribed for her identity, and $\lambda_{x_i}^{\theta_i} = \beta$ otherwise ($x_i \neq \theta_i$). The cost of proposing a link is $c > 0$, and the relation between the parameters in the model is $0 < c < \beta < \alpha$. Note that the cost of proposing a link, c, is paid independently of whether a connection is formed or not.[1]

The main feature of our utility specification is that it captures heterogeneity in several strategic scenarios in a simple way. As a result, we can observe how a player's payoff is affected by the choices of others (i.e., social influence) given her identity. This is motivated by our desire to develop an understanding of how the conflict of preferences, the scenario in which players want to coordinate with others but the preferred choice is not the same for all, interacts in a network game. As discussed in the Introduction, by incorporating players' identities and social influence in the analysis we are extending the applicability of network models to situations in which the preferences of different players may not be aligned.

In order to study the equilibrium of the sequential game, we fix a network con-

[1] We assume c to be lower than β. Otherwise, the only outcome is the empty network because the benefit of coordinating one's behavior to that of a neighbor would not be enough to cover the cost of affiliation. Note that if $\beta < c < \alpha$, this is the case only when choosing the disliked option.

figuration $\{g\}$ generated by the profile $\mathbf{p}$. In the second stage of the game, players decide on an action from the binary choice set X. This is a formal game, represented by $\Gamma = \{N, \{g\}_{i,j\in N}, X, \{\theta_i\}_{i\in N}, \{u_i\}_{i\in N}\}$, and the proper equilibrium concept is the Nash equilibrium. Hence, fix $\{g\}$, a unilateral deviation by player i changes her choice x_i to choice x'_i, where $x_i \neq x'_i$. When no player has incentives to deviate from an action profile $(x_1^*, \ldots, x_n^*)$, it is a Nash equilibrium. Formally:

$$u_i(\theta_i, \mathbf{p}, x_1^*, \ldots, x_i^*, \ldots x_n^*) \geq u_i(\theta_i, \mathbf{p}, x_1^*, \ldots, x'_i, \ldots, x_n^*) \quad \forall\ x'_i \neq x_i^*, \quad \forall i \in N.$$

Note that $u_i(\theta_i, \mathbf{p}, x_1^*, \ldots x_n^*) = u_i(\theta_i, \mathbf{p}, x_i, x_{k_i}(g))$, the actions of players that are not i's neighbors do not change her payoff. The next subsection gives an illustration of the particularities of games with strategic complementarities when identities are introduced in the model. This illustration is represented with the 2-person game, in which the link between the two players is already formed. Thus it only relates to the second stage of the game. Since the cost element in the utility function is common, independently of the adopted behavior the link is already there, we can omit it.

A. The 2-person game

Definition 1. ***Strategic Complements:*** *Let SC be a 2-person game where every player has an identity $\theta_i \in \{0,1\}$ and the finite set of actions X. The payoff matrix, where $2\beta > \alpha > \beta > 0$, depends on each player's choices and identity as follows:*

TABLE 1—PAYOFF MATRICES FOR SC GAMES WITH IDENTITIES.

1 \ **0**	1	0
1	$2\alpha, 2\beta$	α, α
0	β, β	$2\beta, 2\alpha$

$\theta_1 = 1; \theta_2 = 0$

1 \ **1**	1	0
1	$2\alpha, 2\alpha$	α, β
0	β, α	$2\beta, 2\beta$

$\theta_1 = 1; \theta_2 = 1$

0 \ **0**	1	0
1	$2\beta, 2\beta$	β, α
0	α, β	$2\alpha, 2\alpha$

$\theta_1 = 0; \theta_2 = 0$

Each 2×2 coordination game can be played between two players of equal or opposite identities. There are two Nash equilibria in pure strategies and one in mixed strategies.[2] Let us first discuss the pure strategy equilibria. The Nash equilibria (NE) in pure strategies $NE = \{(0,0),(1,1)\}$ present conflicting preferences if players have opposite identities, given each likes a different action but both want to coordinate. Thus, it is not possible to Pareto rank them. However, in games between players with the same identity there is no conflict in preferences, because each one likes the same action (i.e., their behavioral prescription is the same). The equilibrium when both choose the action corresponding to their identity is Pareto dominant in payoffs: $(1,1)$ Pareto dominates $(0,0)$ if two players with identity 1 are playing, and the opposite for two players with identity 0.

In the mixed strategy equilibrium, the probability of choosing one's favorite action when facing a player with the same identity is given by $\underline{q} = (2\beta - \alpha)/(\alpha + \beta)$. When playing against a player with different identity, the result is $\overline{q} = (2\alpha - \beta)/(\alpha + \beta)$. Following (Morris 2000) and (López-Pintado 2006), these probabilities can be understood as the adoption threshold functions, i.e., the influence required from others, the proportion of neighbors making a given choice, so that a player adopts that same action. The inclusion of identities in the analysis gives a new insight to social influence (i.e, threshold models, see (Granovetter 1978)), showing that the q needed varies depending on the identity of the player choosing, but not on the identity of the player(s) she is interacting with. That is, there exist $\underline{q} < \overline{q}$, where $\underline{q}$ is the probability of choosing the liked action and $\overline{q}$ the disliked action. The intuition of this result relates directly to the threshold functions that define the best responses in the Nash equilibrium configurations of the network games. It is also associated to many social scenarios where the utility of affiliation is based on choices of others (i.e., social influence) and not on other's preferences (i.e., identities), but the utility of the individual is based both on her choice and her identity.

III. Equilibrium characterization

In this section we provide the equilibrium characterization for our network game. First we present a categorization of all the possible network configurations that can emerge for any distribution of identities, level of connectivity and action profile chosen. Based on these categories we characterize the set of Nash equilibria for our network game, NE(Γ). That is, the action profile chosen once the network is realized, after the links are proposed. To do this we follow (Hernández, Muñoz-Herrera and Sánchez 2013), who model network games in fixed networks. We extend their analysis with the characterization of the subgame perfect Nash equilibria of the two stage network game. Finally we conclude with a discussion on equilibrium selection.

Notice that along the analysis we can assume without loss of generality a normalization of the utility function for which the cost of link proposal is equal to

[2]We consider a payoff structure such that a player prefers to coordinate in the disliked option than staying alone. This payoff structure is observed in the game of the Battle of Sexes (BOS). For an example of the BOS n-person game see (Zhao, Szilagyi and Szidarovszky 2008).

zero, given the cost of proposal is independent of the action played in the second stage.[3]

A. Network categorization

A player in the network game chooses a vector of link proposals and an action from the set $X = \{0, 1\}$, the same for all her formed connections. The action profiles in the network are such that either all players coordinate on one action (*specialized*) or both actions are chosen by different players (*hybrid*). Given the identity of the players, there are two possible categories, depending on whether all players coordinate in choosing the action they prefer (*satisfactory*) or at least one player chooses the disliked action (*frustrated*).[4] Thus, there are four possible configurations: (i) satisfactory specialized (S_S) where all players coordinate on the same action, which is their preferred choice; (ii) frustrated specialized (F_S), where all players coordinate on the same action, but at least one of them is choosing her disliked option; (iii) satisfactory hybrid (S_H), where all players choose the action they prefer but there is at least one player with a different identity from the rest, so that both actions are present; and (iv) frustrated hybrid (F_H) which portray both actions and at least one player chooses her disliked option. Figure 1 illustrates these categories.

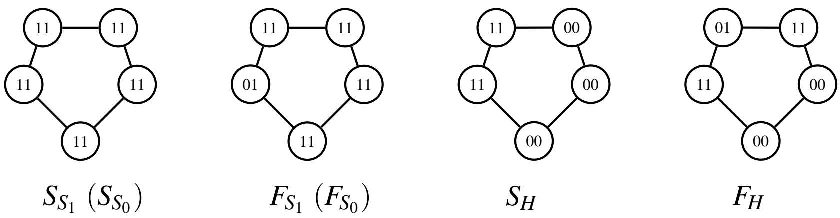


Figure 1. Categories of network configurations. The first digit of a node refers to the identity of a player and the second to her adopted behavior.

B. Nash equilibrium

As mentioned above, given the cost of link proposal is omitted from the best response characterization, once the network is realized, the results in (Hernández,

[3]Once the network is realized, for the computation of the best responses for any player, it affects in the same way the cost of links independently of the action chosen: $[u_i(1, p^i, 1, x_{N_i}(g)) - cp^i] - [u_i(1, p^i, 0, x_{N_i}(g)) - cp^i] = u_i(1, 1, x_{N_i}(g)) - u_i(1, 0, x_{N_i}(g))$. Therefore, this cost is cancelled on both sides of the computation.

[4]We differentiate action profiles as satisfactory or frustrated following the arguments in (Akerlof and Kranton 2000). When a player adopts the behavior prescribed for her identity, this reinforces who she is. However, anyone who chooses the non-prescribed behavior suffers a loss in her identity, entailing a reduction in her utility. That is the reason why $\alpha > \beta$.

Muñoz-Herrera and Sánchez 2013) for fixed networks are applicable to our case. In summary, their findings show that there are two threshold functions in network games with SC when players have conflicting preferences. These thresholds represent the social influence a player's neighbors must exert for her to choose one behavior or another. A function $\underline{\tau}(k_i)$ that represents the minimum number of i's neighbors choosing the action she likes, for her to choose her favorite action as a best response. The threshold function $\overline{\tau}(k_i)$ is the maximum number of neighbors choosing the non-favorite action so that i's best response is still to adopt the behavior she likes, so that if one more of her neighbors chooses the non-favorite action, player i's best response is to change and adopt her disliked option. The results are presented in Proposition 1, where the number of i's neighbors choosing action 1 is χ_i and the number of her neighbors choosing action 0 is $k_i - \chi_i$.[5]

Proposition 1. ***(Hernández et al., 2013)*** *For an SC game, let*

$$\underline{\tau}(k_i) = \lceil \frac{\beta}{\alpha+\beta} k_i - \frac{\alpha-\beta}{\alpha+\beta} \rceil, \tag{2}$$

$$\overline{\tau}(k_i) = \lfloor \frac{\alpha}{\alpha+\beta} k_i + \frac{\alpha-\beta}{\alpha+\beta} \rfloor, \tag{3}$$

defined for any degree $k_i \in \{1, \ldots, n-1\}$. The best response of player i with identity $\theta_i = 1$ and degree k_i, x_i^*, is

$$x_i^* = \begin{cases} 1, & \textit{iff } \chi_i \geq \underline{\tau}(k_i), \\ 0, & \textit{otherwise.} \end{cases} \tag{4}$$

The best response of player i with identity $\theta_i = 0$ and degree k_i, x_i^*, is

$$x_i^* = \begin{cases} 0, & \textit{iff } \chi_i \leq \overline{\tau}(k_i), \\ 1, & \textit{otherwise.} \end{cases} \tag{5}$$

The intuition behind Proposition 1 is that in SC a player i wants to coordinate with the highest number of neighbors making the same choice, and prefers coordination on the action prescribed for her identity. Players with identity $\theta_i = 1$ have incentives to choose the action they like when $\chi_i \geq \underline{\tau}(k_i)$. Thus, players with identity $\theta_i = 0$ choose $x_i = 0$ if $\chi_i \leq \overline{\tau}(k_i)$. Clearly $\overline{\tau}(k_i) > \underline{\tau}(k_i)$ for any k_i, so that a player i requires less influence from her social network to choose what she prefers and more social pressure to adopt her disliked behavior, compared to an analysis ignoring identities. For instance, returning to the example of people choosing between two technologies, say two operative systems such as MacOS and Windows, those who prefer Mac over Microsoft need less support from their

[5]Denote by $\lceil \ldots \rceil$ and $\lfloor \ldots \rfloor$ respectively the maximum lower integer or the minimum higher integer of the real number considered.

friends to purchase this operative system. However, they would require more pressure from their friends to buy the Windows system that they dislike. The two tipping points are illustrated in Figure 2.

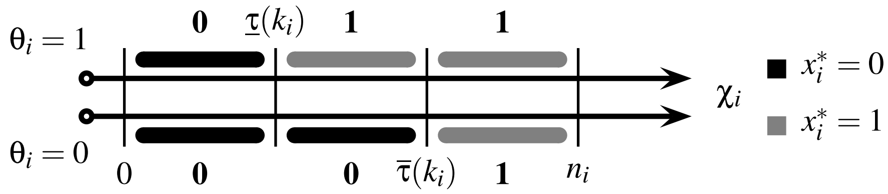


Figure 2. SC Adoption Thresholds

From the best responses characterized above, it is clear that there will be very many different equilibria. For an illustration consider the following case: Satisfactory specialized equilibria (S_S) are a very restrictive case in which all players must have the same identity. Assume $\theta_i = 1$ for all $i \in N$. Then, in a S_S equilibrium, all players choose their preferred behavior; action 1. However, if for any reason a player or group of players chose action 0 and for them the condition $\chi_i \geq \underline{\tau}(k_i)$ is not verified, a frustrated Nash equilibrium emerges, i.e., *frustrated hybrid*. This because all players have the same identity but some are choosing the disliked option. In general, when all players share a common identity, if an equilibrium is satisfactory it has to be specialized. There is another manner in which specialized equilibria emerge, namely when the distribution of identities is not homogeneous but both social categories are present (there are 1's and 0's) and either condition $\chi_i \leq \underline{\tau}(k_i)$ or $\chi_i \geq \overline{\tau}(k_i)$ holds for all players. As a consequence, for the same distributions of links and identities, two players with opposite identities can best respond with the same action and vice versa. This points to conditions where social pressure can exert more influence than the individual preference of a player who is interacting with others, when choosing what behavior to adopt. Examples of Nash equilibria are illustrated in Figure 3.

C. *Subgame perfect Nash equilibrium*

Our analysis so far has focused on the best response when players play once the network is formed. We now proceed to the first stage of the network game: *affiliation*. By backward induction analysis we develop a characterization of the subgame perfect Nash equilibria (SPNE). In our case, all players play simultaneously at each stage. Thus, we are interested in knowing which vector of link proposals is part of an equilibrium. Notice that a given network can be generated from different vectors of link proposals. For instance, if player i has k_i neighbors in $\{g\}$, it could be because she proposed a link to only her k_i neighbors, or because she proposed links to those and even more players; who did not proposed

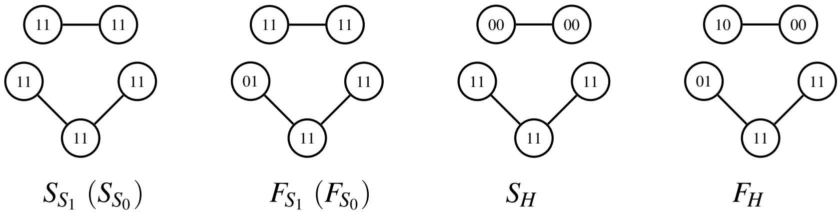


Figure 3. Examples of Nash equilibria. The first digit of a node refers to the identity of a player and the second to her adopted behavior.

a link back to i. The first Lemma states that in a SPNE for a given network $\{g\}$ the best proposed links $\mathbf{p^i}$ for any player i does not exceed the set of her realized neighbors in $\{g\}$.

Lemma 1. *Let $\{g\}$ be a network where player i has k_i neighbors denoted by $\{i_1, i_2, \ldots, i_{k_i}\}$. Consider two vectors of proposals:*

- $\mathbf{p^i}$ *with* $p^i_j = 1$ *if* $j \in \{i_1, i_2, \ldots i_{k_i}\}$ *and* $p^i_j = 0$ *if* $j \notin \{i_1, i_2, \ldots i_{k_i}\}$
- $\tilde{\mathbf{p}}^\mathbf{i}$ *with* $\tilde{p}^i_j = 1$ *if* $j \in \{i_1, i_2, \ldots i_{k_i}, z_i, z_2, \ldots z_s\}$ *and* $\tilde{p}^i_j = 0$ *if* $j \notin \{i_1, i_2, \ldots i_{k_i}, z_i, z_2, \ldots z_s\}$.

For the game Γ where $\{g_{ij}\}_{i,j\in N}$ is realized then

$$u_i(\theta_i, \mathbf{p^i}, x^*_1, \ldots, x^*_i, \ldots x^*_n) \geq u_i(\theta_i, \tilde{\mathbf{p}}^\mathbf{i}, x^*_1, \ldots, x^*_i, \ldots x^*_n)$$

where the set of players $\{z_1, z_2, \ldots, z_s\} \cap \{i_1, i_2, \ldots i_{k_i}\} = \emptyset$.

Proof: It is straightforward to check that

$$u_i(\theta_i, \tilde{\mathbf{p}}^\mathbf{i}, x^*_1, \ldots, x^*_i, \ldots x^*_n) = u_i(\theta_i, \mathbf{p^i}, x^*_1, \ldots, x^*_i, \ldots x^*_n) - c|\{z_1, z_2, \ldots, z_s\}|$$

As a consequence of the above Lemma, networks where players proposed the final links will be the survival networks in the backward induction process. In particular, a SPNE is a network that results given that in the *affiliation* stage no link proposal is unreciprocated, and in the *behavior adoption* stage players choose according to Proposition 1. Nonetheless, the analysis of subgame perfection does not permit us to discriminate enough, and there are multiple surviving configurations that satisfy these conditions. In the last part of this section, we will refine the analysis of equilibria by considering some selection criteria that have proven to be essential for network studies.

D. Equilibrium selection

To model equilibrium selection, we use two different concepts that are commonly applied to network games: Pairwise stability (Jackson and Wolinsky 1996) and efficiency (i.e., utilitarian welfare). Our aim is to discriminate equilibria in terms of how they dominate in payoffs and how likely is it for players to be satisfied and adopt the behavior prescribed for their identities in the presence of social influence from their neighbors.

We begin by evaluating for which networks, once an action profile is chosen, players have incentives to change their connections (i.e, increase or decrease their degree). Because pairwise stability only takes into account link selection, we fix the set of action profiles, x, to establish what a stable network is. The original concept of pairwise stability, by (Jackson and Wolinsky 1996), states that a network is pairwise stable with respect to the total value of the network (often the aggregate utility of all nodes) and an allocation rule (how that utility is divided among nodes) if (i) there is no player who is better off by unilaterally cutting one of her existing links and (ii) if there is no pair of unconnected players who would benefit from creating a link between them; if one of them is better off by forming the link then the other is worse off by doing so[6]. We adapt and formally the define the concept for our model as follows:

Definition 2. ***Pairwise stability:*** *Let x be an action profile. A network $\{g\}$ generated by $\mathbf{p}$ is pairwise stable if:*

1) Suppose $p^i_j = p^j_i = 1$. Consider the network $\tilde{g}$ generated by $\tilde{\mathbf{p}}$ that coincides with g except $\tilde{p}^i_j = \tilde{p}^j_i = 0$, (i.e., players i and j are not connected) then

$$u_i(\theta_i, \mathbf{p^i}, x) \geq u_i(\theta_i, \tilde{\mathbf{p}}^\mathbf{i}, x) \text{ and } u_j(\theta_j, \mathbf{p^j}, x) \geq u_j(\theta_j, \tilde{\mathbf{p}}^\mathbf{j}, x)$$

2) Suppose $p^i_j = p^j_i = 0$. Consider the network $\tilde{g}$ generated by $\tilde{\mathbf{p}}$ that coincides with g except $\tilde{p}^i_j = \tilde{p}^j_i = 1$, (i.e., players i and j are connected) then

(a) if $u_i(\theta_i, \mathbf{p^i}, x) < u_i(\theta_i, \tilde{\mathbf{p}}^\mathbf{i}, x)$ then $u_j(\theta_j, \mathbf{p^j}, x) > u_j(\theta_j, \tilde{\mathbf{p}}^\mathbf{j}, x^*)$ or

(b) if $u_j(\theta_j, \mathbf{p^j}, x) < u_j(\theta_j, \tilde{\mathbf{p}}^\mathbf{j}, x)$ then $u_i(\theta_i, \mathbf{p^i}, x) > u_i(\theta_i, \tilde{\mathbf{p}}^\mathbf{i}, x)$

Since in our model players choose both links and actions, we provide now a definition of pairwise stable networks in our game, in order to take into account not only links selection but also the action profile, x, chosen at Stage 2. To do so, we integrate the subgame perfect framework with conditions of an adapted concept of pairwise stability, and evaluate how increasing or decreasing the density of the network affects players' payoffs. Therefore, we select the set of action profiles to those leading to Nash equilibrium in the second stage which correspond to a Nash Equilibrium profile even when the network configuration changes. The following definition captures this idea:

Definition 3. *The pair $(\{g\}, x^*)$ is a pairwise stable Nash equilibrium if*

[6] For different theoretical characterizations of pairwise stability see also (Jackson and Watts 2001) and (Calvó-Armengol and İlkılıç 2009).

1) g is pairwise stable, and

2) For the network $\tilde{g}$ generated by $\tilde{\mathbf{p}}$ that coincides with g except $\tilde{p}^i_j = \tilde{p}^j_i = 0$ or $\tilde{p}^i_j = \tilde{p}^j_i = 1$ the action profile x^* is a Nash equilibrium in g and $\tilde{g}$.

The next lemma characterizes the set of pairs $(\{g\}, x^*)$ which are pairwise stable Nash equilibria in the game Γ. First, it describes the network structures which satisfy the pairwise stability property. Finally, the set of Nash equilibrium action profiles which verify the stability notion.

Lemma 2. *The pair $(\{g\}, x^*)$ is a pairwise stable Nash equilibrium of the game Γ if it satisfies the following conditions:*

*(i) Every player i is connected to **all** other players in the network who are choosing the same action as her: if $x^*_i = x^*_j$ for any $i, j \in N$, then $g_{ij} = 1$, and*

*(ii) Every player i is connected **only** to players who are choosing the same action as her: if $g_{ij} = 1$ for any $i, j \in N$, then $x^*_i = x^*_j$*

(iii) Let be $\theta_i = 1$, and χ_i the number of neighbors of player i playing action 1 in the network g. Then,

a) if $x^*_i = 1$ then $\chi_i \geq \underline{\tau}(k_i + 1)$

b) if $x^*_i = 1$ then $\chi_i - 1 \geq \underline{\tau}(k_i - 1)$

c) if $x^*_i = 0$ then $\chi_i + 1 < \underline{\tau}(k_i + 1)$

d) if $x^*_i = 0$ then $\chi_i < \underline{\tau}(k_i - 1)$

The conditions for players with $\theta_i = 0$ are symmetric.

Proof:

From this point on, and abusing notation, we will use $\{g\}$ and $\mathbf{p}$ indistinctively in the utility function, given each $\mathbf{p}$ generates a unique $\{g\}$.

Let us prove that a network structure produces a pairwise stable Nash equilibrium if every player is connected to *all* others coordinating their behavior with her. Consider two networks:

- $\{g\}$ where $x^*_i = 1$ for player i, and there is at least one player j choosing $x^*_j = 1$, such that $g_{ij} = 0$, and
- $\{\tilde{g}\} \supset \{g\}$ in which i and j form a link between them, $\{\tilde{g}\} = \{g\} + g_{ij}$

For the game Γ:

$$u_i(\theta_i, \{\tilde{g}\}, x^*_1, \ldots, x^*_i, \ldots x^*_n) > u_i(\theta_i, \{g\}, x^*_1, \ldots, x^*_i, \ldots x^*_n), \text{and}$$

$$u_j(\theta_j, \{\tilde{g}\}, x^*_1, \ldots, x^*_j, \ldots x^*_n) > u_j(\theta_j, \{g\}, x^*_1, \ldots, x^*_j, \ldots x^*_n)$$

where it is straight forward to check that

$$u_i(\theta_i, \{\tilde{g}\}, x_1, \dots, x_n) = \alpha(k_i + 1) - ck_i > \alpha k_i - c(k_i - 1) = u_i(\theta_i, \{g\}, x_1, \dots, x_n)$$

since $\alpha > c$. From this, it derives that if player i is linked to k neighbors, her utility is increasing in k as long as they choose her same action.

We show now that a network forms a pairwise stable Nash equilibrium if every player is connected *only* to neighbors coordinating their behavior with her. Consider two networks:

- $\{g\}$ where $x_i^* = 1$ for player i, and $\chi_i < k_i$ of i's neighbors play $x_j^* = 1$, while $(k_i - \chi_i) > 0$ play $x_j^* = 0$, and
- $\{\tilde{g}\} \subset \{g\}$ in which i drops any neighbor j whose action is $x_j^* = 0$.

For the game Γ:

$$u_i(\theta_i, \{\tilde{g}\}, x_1^*, \dots, x_i^*, \dots x_n^*) > u_i(\theta_i, \{g\}, x_1^*, \dots, x_i^*, \dots x_n^*)$$

It is straightforward to check that

$$u_i(\theta_i, \{\tilde{g}\}, x_1^*, \dots, x_i^*, \dots x_n^*) = u_i(\theta_i, \{g\}, x_1^*, \dots, x_i^*, \dots x_n^*) + cI_{\{x_j \neq x_i\}}$$

Finally, the lemma states the conditions under which the equilibrium action profile x^* is *robust* to the addition or the removal of one link (i.e. it is not profitable for any player to change her action after adding any possible new neighbor or removing an existing one, regardless of the behavior she may adopt). Then the conditions relating χ_i and the threshold functions from (Hernández, Muñoz-Herrera and Sánchez 2013) must be satisfied when a player's degree is increased or decreased by 1.

Given the utility structure of player i, when two players coordinate in forming a link between them but their adopted behavior is uncoordinated, say i chooses $x_i = 1$ and j chooses $x_j = 0$, there is no positive payoff for any of them from this relationship. On the contrary, there is a negative payoff in terms of the cost of relating, without the complementarities from choosing the same action. Therefore, players prefer networks where everybody in their neighborhood plays the same action they play, and any link to a neighbor who is behaving differently is eliminated. The intuition behind Lemma 2 points to a single argument: for each action profile x^* there is only one network configuration which conforms a pairwise stable Nash equilibrium.

Up until now we have defined and analyzed pairwise stability and pairwise stable Nash equilibria for our network game. For the remaining part of this section we introduce a new concept for equilibrium selection: efficiency.

Let the value of a pair $(\{g\}, x)$ be the aggregate of individual utilities:

$$v(\{g\}, x) = \sum_{i=1}^{n} u_i(\theta_i, \mathbf{p}, x_i, x_{k_i(g)})$$

From this, it follows that a pair $(\{g\}, x)$ is *efficient* if $v(\{g\}, x) \geq v(\{\tilde{g}\}, \tilde{x})$, $\forall \{g\} \neq \{\tilde{g}\}$ and $\forall x \neq \tilde{x}$. The next definition formally expresses the idea:

Definition 4. ***Strong Efficiency:*** *A pair* $(\{g\}, x)$ *is strongly efficient in the game* Γ *if* $(\{g\}, x) = \underset{\{g\},x}{\text{argmax}}\ v(\{g\}, x)$.

We derive from the concept of pairwise stability that only two kind of network configurations can conform a pairwise stable Nash equilibrium: (1) a completely connected structure if the action profile is *specialized*, and (2) a network with two isolated and completely intra-connected components if the action profile is *hybrid*, where each component is specialized in a different action. For an illustration see Figure 4.

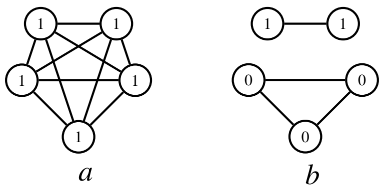


Figure 4. Pairwise stable Nash equilibria. Network *a* portrays the specialized case for action 1. Network *b* portrays the hybrid case. The digit of a node refers a player's adopted behavior.

Notice that pairwise stability is absent of the inclusion of identities, and just tackles the selection of structures by ranking one configuration over another if both have the same action profile. To take identities into account, we discuss next the equilibrium selection for different compositions of the population of players. As a consequence, this characterization depends on the *a priori* distribution of identities. We will refer to the distribution of identities, the share of players with identity 1 or 0, as the indicator for the *level of conflict in preferences* in the game, denoted by Φ. This will be particularly useful in our experimental study, presented in the next section. We assume there is a proportion of θ_i players with identity θ_i, where $\theta_0 + \theta_1 = 1$. Using the share of players with identity 1 as the reference group, we define the level of conflict in preferences as the binary entropy function of the distribution of identities, where $\Phi \in (0, 1)$. The more homogeneous a population is, the lower the level of conflict. Thus, if $\theta_1 = 0$ or $\theta_1 = 1$, then $\Phi = 0$. The more heterogeneous the population is the higher the level of conflict. This means that if $\theta_1 = \theta_0$, then $\Phi = 1$. See Figure 5 for an illustration.

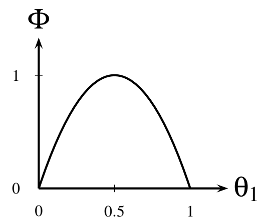


FIGURE 5. THE HORIZONTAL AXIS REPRESENTS THE SHARE OF PLAYERS WITH IDENTITY 1 (θ_1) IN THE POPULATION. THE VERTICAL AXIS REPRESENTS THE LEVEL OF CONFLICT (Φ).

Based on this consideration of conflicting preferences, we want to know what are the conditions, in terms of Φ, for players to choose a specialized or a hybrid equilibrium. In order to achieve this we characterize efficiency as a measure of utilitarian welfare and denote a network as efficient if it maximizes the aggregate utility of all players, given the distribution of identities (i.e., the level of conflict in preferences). Lemma 3 presents this arguments.

Lemma 3. *The strongly efficient configuration of the game Γ is the complete network specialized in the prescribed behavior for the majority, for any level of conflict. So that:*

(i) if $\Phi = 0$, such that $\theta_1 = 1(0)$, $x_i^* = 1(0)$ and $k_i = (n-1)\ \forall\ i \in N$

(ii) if $0 < \Phi < 1$, such that $\theta_1 > \theta_0 > 0$, $x_i^* = 1$ and $k_i = (n-1)\ \forall\ i \in N$

(iii) if $\Phi = 1$, such that $\theta_1 = \theta_0$, either $x_i^* = 1$ or $x_i^* = 0$, and $k_i = (n-1)\ \forall\ i \in N$

Proof: The first element follows because a network in which all players who adopt the same behavior are affiliated dominates in payoffs any less connected network. Moreover, such a network will rank the highest if all players are choosing the behavior they like. For the case of $\Phi = 0$ this is the Satisfactory Specialized (S_S) configuration.

To prove the second element we compare two networks. A satisfactory hybrid configuration (S_H), in which all players choose the action they like, and a frustrated specialized configuration (F_S), in which all players choose the action of the majority. It follows from the statement above that a F_S in the action preferred by the minority will be dominated in payoffs, given $\alpha > \beta$. Also, it follows from Lemma 2 that such networks are pairwise stable, so that $k_i = n-1$ for all players in the F_S, and $k_i = n\theta_i$ for players in the S_H.

Consider a distribution of identities such that there are $n\theta_1$ players with identity 1 and $n\theta_0 = n(1-\theta_1)$ players with identity 0. The aggregate payoffs of the F_S network are given by:

$$v(F_S) = \sum_{i=1}^{\theta_1 n} \alpha n + c(n-1) + \sum_{i=1}^{\theta_0 n} \beta n + c(n-1) \tag{6}$$
$$= n[\theta_1(\alpha n - c(n-1)) + (1-\theta_1)(\beta n - c(n-1))]$$

The aggregate payoffs of the S_H network are given by:

(7)
$$v(S_H) = \sum_{i=1}^{\theta_1 n} \alpha(\theta_1 n) + c(\theta_1(n-1)) + \sum_{i=1}^{\theta_0 n} \alpha\theta_0 n + c(\theta_0(n-1))$$
$$= n[\theta_1(\alpha\theta_1 n - c(\theta_1(n-1))) + (1-\theta_1)(\alpha(n-\theta_1 n) - c(n-\theta_1(n-1)))]$$

where it is straightforward to check that

$$v(F_S) > v(S_H) \text{ for } \theta_1 \geq \frac{1}{2} \tag{8}$$

The third point is easy to prove under the conditions exposed so far, because if $\Phi = 1$ then $\theta_1 = \frac{1}{2}$, and Equation 8 states that under that level of θ_1 the aggregate generated profit in a Frustrated Specialized configuration is higher than in a Satisfactory Hybrid one. Obviously the profit is the same if the action chosen by all players is 0 or 1, since $\theta_0 = \theta_1 = \frac{1}{2}$.

The intuition of this Lemma is that if all players behave as prescribed for the majority, so that the entire population is specialized, this is strictly better from a social welfare perspective, than if each player adopts her preferred behavior and the population segregates. That is, when social influence in the population is exerted by a majority, it is socially better for the members of the minority to choose the behavior that goes against their identity but increases their benefits from the complementarities of their neighbors. Specialization is socially better even if the share of each social category in the population is exactly divided into equal parts. Particularly, when this is the case, socially there is no difference on which behavior players specialize in. Both specializing in 1 or 0 gives the same aggregate value. Nonetheless, this is the aggregate welfare and for cases with a strict majority it is not always the case that the minority maximizes individual payoffs by following this strategy. In fact, a player i from the minority gets higher payoffs in the satisfactory specialized network in which each component is completely connected as long as $\theta_i > \frac{(\beta-c)}{(\alpha-c)} > \frac{1}{2}\frac{(\alpha-\beta)}{(\alpha-c)}$.[7]

Going back to our example on the adoption of technologies we have that a network is pairwise stable if all players purchasing MacOS are connected and none of them relates to anyone purchasing Windows, and viceverza. Furthermore, if those who like MacOS more than Windows are a majority, it is better for everyone in the society to buy this operative system, even for those who like Windows. By doing so they can all relate between each other and obtain greater benefits from the compatibility of their choices than if they had segregated into clusters of Mac users and Windows users.

Finally, note that there is an important consideration when relating efficiency and pairwise stability. If the satisfactory hybrid equilibrium emerges, so that players are segregated by identities in two components each choosing the preferred action of the players in the component, there is no smooth transition to the specialized frustrated (efficient) configuration. Once a player has entered a pairwise stable but non efficient network, there are no individual incentives to move to the efficient one. A player who is part of the majority has only incentives to link to a player from the minority if she knows the other will choose her frustrated action. A player who is part of the minority has no incentives to unilaterally or bilaterally deviate to the component where the majority is segregated, because she would need multiple changes to be connected to all of them. Such a transition requires a stronger restriction than *dyadic* coalitions as modeled in pairwise stability.

In conclusion, our identity-based model and equilibrium characterization points to the following considerations. First of all, players are always better off coordinating with all their neighbors in the same behavior, because social influence from others results in greater benefits from the complementarities of the interaction. If this is not the case, a player will rather eliminate a relationship with an *uncoordinated* neighbor. However, given the interaction between identity-based payoffs and the effect of others' behavior on a player's choices, it is not always the case that players adopt the behavior prescribed for their identity. Our model of conflicting preferences shows that depending on the distribution of identities (i.e., the level of conflict in the population) some equilibria dominate others. In particular, equilibria in which all players are integrated into one same component and the behavior prescribed for the majority is the specialized action are the socially efficient networks. Moreover, in many cases, as long as the size of the minority is not big enough, these *frustrated* equilibria are also dominant on individual payoffs. Notice that in this order of ideas, the share of the population that a given identity occupies can determine whether players belonging to this group will be governed by social pressure and sacrifice their identity-based preferences for their social interaction benefit. In the next section we describe the experimental study we have used to test our game theoretic model.

[7] This comes from the comparison of choosing the behavior of the majority in the complete network or the preferred choice in the network segregated into two components: $\alpha n\theta_i - c(\theta_i n-1) \geq \beta n - c(n-1)$.

IV. The experiment

Our theoretical model gives account of the way equilibrium takes place in network games in which identities and social influence are at play. To test the results of our theory we designed an experimental game which replicates our identity-based model in the laboratory. Our interest is to evaluate the interplay between individual preferences and social influence by assessing the effect that different levels of conflict in preferences have on individual and aggregate behavior. As our game-theoretic analysis shows, there are multiple configurations in equilibrium that are likely to emerge and their likelihood depends on the strength of individual identities and on the influence exerted by others.

A. The experimental game

There are 15 subjects in a one-shot network game interaction. Each subject at the beginning of the interaction is informed about a symbol she is assigned to, either a *square* or a *circle*. The two symbols represent the artificially generated social categories to which subjects can belong to. Participants were also informed of how many of the remaining 14 subjects in the population had been assigned to each category (how many were circles and how many were squares).

The experimental game replicates the two-stage structure of our game theoretic model. In the first stage, *affiliation*, subjects simultaneously decided to whom in the group of 15 subjects they wanted to propose a link to (see Figure 6). Subjects were also assigned an identification number from 1 to 15 to facilitate the linking process. The identification numbers were randomly associated to the social categories but kept the same for all groups (i.e., subject with identification number 12 always belonged to the social category square). The cost of proposing a link is $c = 2$ and, only if two subjects proposed to each other a connection between them was created.

In the second stage, subjects were informed about the proposals made and connections formed in their group (see Figure 7). That is, subjects were informed of the social network that resulted from the affiliation stage. Then, they had to choose an action *up* or *down*. Up (down) gives $\alpha = 6$ points to a subject with identity *square* (*circle*) for every neighbor she coordinates with in the same choice. Down (up) gives her $\beta = 4$ points. These choices represent the identity-prescribed behavior from our model.

The total number of points earned is calculated with the payoff function in Equation 1. This linear payoff function makes it straight forward for participants to calculate their expected payoffs in any situation given their individual preferences and the behavior of others (i.e., social influence). In addition, all subjects received a printed table illustrating the points they can get for any level of connections (from 1 to 14) and any choice in which they coordinated on (up or down).

Therefore, in the experiment subjects are assigned to one of two social cate-

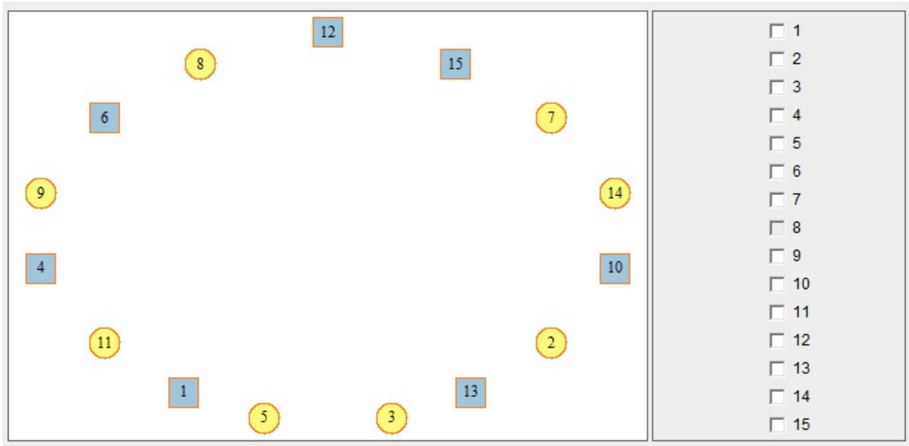


FIGURE 6. SCREEN OF CONNECTION PROPOSALS.

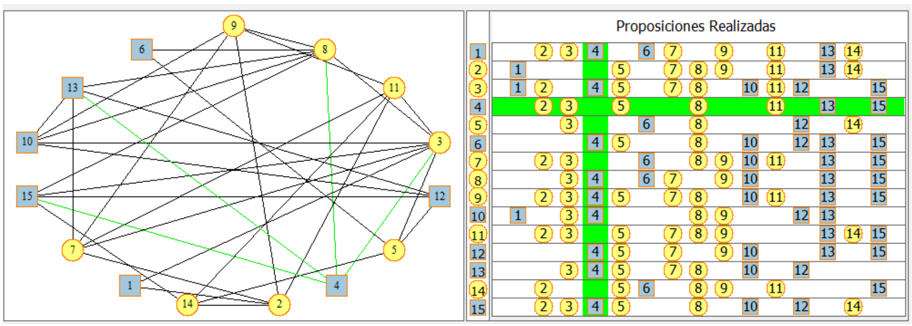


FIGURE 7. SCREEN WITH RESULTING NETWORK AND PROPOSALS MADE.

gories. Each of the categories has a behavior that gives more benefits if chosen (in isolation) by a subject belonging to that category. Subjects, knowing their own identity and that of all other participants, propose to establish relationships between them. Only if proposals are mutual a network connection is formed. Once they observe the resulting network, subjects simultaneously choose an action (i.e., adopt a behavior). For every connected neighbor with whom a subject coordinates in her behavior, she benefits as a consequence of the strategic complementarities of the interaction. If two subjects are connected but choose differently, they do not benefit from the relationship but have to pay the cost of affiliation. If two subjects are not connected, they cannot influence each other, because their choices have no effect on the other's payoffs. In this way our experimental game addresses the question of how individual preferences and social influence affect the decision-making process of whom to relate with and what behavior to adopt.

B. *Experimental design and treatments*

The experiment consists of a two-stage network game played for 25 periods by groups of 15 subjects. It was conducted in the Laboratory of Experimental Economics (LINEEX) at the University of Valencia in November 2012. Subjects interacted through computer terminals and the experiment was programmed using z-Tree (Fischbacher 2007).

Upon arrival subjects drew a ballot to be randomly assigned to a seat in the laboratory. At the beginning of the experiment instructions were read out loud to all subjects to guarantee that they all received the same information (you can see a full copy of the instructions in appendix A). Instructions also appeared on their screens. 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. In every period subjects were randomly matched using a strangers protocol, so that each round represented an independent one-shot interaction with no reputation effects. Identities were randomly assigned in the first round and kept constant along the 25 interactions and only group composition varied (also the assigned identification number varied). That is, the social category a subject belonged to was always the same for all rounds. The first five periods were trial rounds.

To evaluate the effect that the level of conflict in preferences has upon outcomes, the interplay between individual preferences and social influence, we used the distribution of identities in the groups as our experimental variable. We implemented three treatments that systematically vary this feature: No conflict, Low conflict and High conflict (see Table 2). In all our treatments we kept the social category *square* to be the majority. Therefore, for all treatments the socially efficient and individual payoff dominant outcome is the complete network specialized in choosing *up*, the prescribed behavior for the majority.

Our experimental design captures an important mixed-motive social situation derived from our theory, which results from the incorporation of identities into the

TABLE 2— NUMBER OF SUBJECTS IN THE MAJORITY AND THE MINORITY PER TREATMENT, AND SIZE OF THE SUBJECT SAMPLE PER TREATMENT/SESSION.

Treatment	Majority	Minority	Session
No conflict	15	0	30
Low conflict	12	3	45
High conflict	8	7	45

analysis. Subjects earn more by coordinating their choices with others, maximizing payoffs when the entire group integrates and coordinates in one same choice (social motive). This motive results from the effect that social influence has on behavior. But subjects disagree in their preferences about which action to coordinate on (individual motive). This motive results from the effect that individual preferences have on behavior. From this, we derive contrasting hypotheses for the equilibrium selection strategies. On one hand, we use the category motivation in the identity literature stating that if there are artificially induced identities, subjects are more likely to favor their in-group. On the other, we use the payoff dominant motivation from the literature on social influence in network interactions, which is independent of the identities of the players, and states that if subjects can decide with whom to connect they are more likely to coordinate in the equilibrium that gives them the highest payoffs. Thus, hypotheses 1a and 2a, the identity-dominant hypotheses, are a result of how identity predicts equilibrium selection in our game. Hypotheses 1b and 2b, the payoff-dominant hypotheses, are a result of how social influence predicts equilibrium selection for rational payoff maximizers in our game. The hypotheses for the affiliation stage of the experimental game are:

Hypothesis 1a. *(Identity-dominant affiliation) The higher level of conflict the higher the tendency to propose connections only to the in-group.*

Hypothesis 1b. *(Payoff-dominant affiliation) The level of conflict does not affect the tendency to propose connections to the in-group (the same amount of links to in-group and out-group, adjusted for group size).*

The affiliation hypothesis argue that the probability of linking with one's in-group or out-group is the same if subjects aim to maximize payoffs. However, if subjects rather strengthen their social identity, it is more likely to be connected to one's in-group. Therefore, integration between identities is predicted for all treatments by Hypothesis 1b, and segregation is predicted for treatments with conflicting preferences by Hypothesis 1a.

Hypothesis 2a. *(Identity-dominant behavior) The higher the level of conflict the more likely subjects will adopt the behavior they prefer as prescribed by their social category.*

Hypothesis 2b. *(Payoff-dominant behavior) The level of conflict will have no effect and subjects will adopt the behavior preferred by the majority.*

The behavior adoption hypotheses state that if identity is more salient than social influence (i.e., payoffs), the treatments with positive level of conflict are more likely to result in satisfactory hybrid action profiles (each subject chooses the behavior she *prefers*). Otherwise, the frustrated specialized action profile will be the outcome (subjects in the majority choose what they prefer and subjects in the minority choose what they do not prefer). Particularly, for the No Conflict treatment the satisfactory specialized outcome is predicted. Consequently, we use this treatment as our baseline condition.

Finally, we derive point predictions from our theory in relation to our characterization of equilibrium and the selection criteria modeled. Notice, nevertheless, that as argued by (Camerer 2003), it is unlikely that equilibrium is reached instantaneously in one-shot games. It has been pointed along the extensive experimental research on rational behavior that the idea of instant equilibration is so unnatural that perhaps an equilibrium should not be thought of as a prediction which is vulnerable to falsification at all. A more useful perspective should be to perceive equilibrium predictions as the limiting outcome of an unspecified learning process that unfolds over time. This means that we could expect to observe learning from the repetition of the interactions in the experiment. In this view, equilibrium is the end of the story of how strategic thinking, optimization, and equilibration (or learning) work, not the beginning (one-shot) or the middle (equilibration). The following are the hypotheses on equilibrium derived from our game theoretic model:

Hypothesis 3. *(Subgame Perfection) The higher the number of one-shot interactions subjects are part of, the more likely the difference between links proposed and links formed will be reduced.*

This prediction is derived for the affiliation stage of our network game from the backward induction process. Finally, the hypothesis on pairwise stability is derived from our modeling of equilibrium selection:

Hypothesis 4. *(Pairwise stability) The higher the number of one-shot interactions subjects are part of, the more likely subjects choosing the same action will be neighbors.*

From these previous hypotheses we can state that if learning is manifested along the repeated interactions, subjects choosing the same behavior are more likely to be connected, regardless of whether identities or social influence motivate their behavior. Specifically for the payoff dominant strategies networks will be completely connected into a single component, so that the efficient configuration will emerge. Otherwise, the segregated configuration where players are separated into social categories should be observed. Nonetheless, whether it is one or two components, these hypotheses predict that networks will tend to be more dense along time, leading towards the pairwise stable predicted configurations.

C. Experimental procedures, data and methods

All subjects in our experiment were students from the campus of social sciences of the University of Valencia (Spain). Subjects were recruited through online recruitment systems. In total 120 subjects participated in three sessions, one for each treatment (No, Low and High Conflict). There were 30, 45 and 45 participants in each session, respectively. Each session lasted between 90 and 120 minutes and no one participated in more than one session. On average everyone earned 16.5 euros, including a show-up fee of 5 euros.

To conclude this section, we describe the measures we use to test the hypotheses presented above, and the way we developed our analytical strategy. Recall that in reference to a subject, others either belong to her in-group, when they share her identity, or to her out-group, when identities are different.

In-group favoritism. To assess a subject's favoritism to propose connections to the in-group rather than the out-group, the number of proposals sent by a subject to the in-group was divided by the subject's total number of proposals sent. In-group favoritism could range from a maximum of 1 where all proposals were sent to the in-group to a minimum of 0 where all proposals were sent to the out-group. A value of 0.5 denoted equal preferences for sending proposals to both the in-group and the out-group.

Reciprocation. Reciprocation was operationalized as a subject's number of reciprocated proposals (i.e., realized connections) divided by a subject's total number of proposals, regardless of group membership. Reciprocation had a maximum of 1 (0) when all proposals were reciprocated (rejected), and hence no coordination problem occurred in the affliation stage.

Pairwise stability. Pairwise stability was measured with a subject's number of realized connections with in-group members as compared to the total possible connections with this group. That is, the number of in-group members minus the subject. Subjects in the No Conflict condition could realize up to 14 in-group connections, subjects in the Low Conflict condition could realize up to 11 (majority) or 2 (minority) connections, and subjects in the High Conflict condition could realize up to 7 (majority) or 6 (minority) connections. Again, a value of 1 expressed maximum pairwise stability. That is, a subject sent proposals to all of her in-group members of which all proposals were reciprocated, resulting in the subject's connection with every in-group member. Note that this measure disregarded activities with the out-group.

Analytical strategy. The data structure at hand did not permit standard ordinary least square regression modeling. Standard regression models base on the assumption that observations are measured independently from one another. This independence assumption was violated in our data: The experiment included 120 subjects who each played 20 one-shot interactions, so that a total of 2,400 interactions (Level 1) were nested within clusters of 120 subjects (Level 2). Interactions belonging to the same subject could not be assumed to occur independently from one another, as different subjects likely followed varying behavioral tendencies.

For example, throughout all interactions, some subjects may have systematically favored in-group members more than may have other subjects.

Multilevel regression modeling is a methodology for the analysis of complex data patterns with a focus on nesting (Snijders and Bosker 2012). Such models allow variability at multiple levels of observations, namely variability between interactions (Level 1) and variability between subjects (Level 2). While the interpretation of these models is comparable to standard regression models, they additionally assume the intercept (and sometimes the slope) to be randomly varied for each of the 120 subjects. These models, in the following referred to as mixed-models, allowed subjects to differ in their general behavior with regard to in-group favoritism, reciprocation and pairwise stability. Three separate models were run for in-group favoritism, reciprocation and pairwise stability.

V. Results

Our experimental study assesses the interplay between individual preferences and social influence by varying the level of conflict in preferences in a network game with strategic complementarities. In this section we describe our main findings beginning with a descriptive discussion of the behavior of the participants. The data show that nearly all choices corresponded with the subjects' preference. We observed that 99.3 percent of the decisions on behavior adoption where such that the prescribed behavior for the social categories was selected. For the affiliation criteria it was found that 99.4 percent of the connections were formed between subjects choosing the same behavior. Table 3 presents an overview of the proposals sent and reciprocated (i.e., the realized connections) for the different experiment conditions and groups. In-group favoritism and reciprocation were most prominent in the No Conflict condition. Stronger in-group favoritism related to increased reciprocation (Pearson's correlation coefficient: $r = .694, p < .001$), which in turn was associated with greater pairwise stability ($r = .687, p < .001$).

Hypothesis 1a (Identity-dominant affiliation) expected that higher level of conflict would lead to greater favoritism for in-group proposals. The alternative *Hypothesis 1b (Payoff-dominant affiliation)* stated that no such effect would occur. Table 4 presents the results from the mixed-effects regression models. The constant of 0.90 indicates that in-group favoritism was generally high: put aside all other variables (experimental conditions, group membership and development over periods), it could be predicted that subjects send proposals to members from their own group in 90 percent of the cases. According to the negative and significant parameter estimate in Model A, subjects in the Low Conflict condition showed less in-group favoritism than subjects in the No Conflict condition (which served as the reference category). Subjects in the High Conflict condition did not differ significantly in their favoritism from the No Conflict group. This suggests that in-group favoritism was greater in the High Conflict group than in the Low Conflict group, supporting *Hypothesis 1a* over *Hypothesis 1b*.

Hypotheses on behavior adoption stated that if subjects were more influenced by

TABLE 3—SUBJECTS' PROPOSALS AND CONNECTIONS WITHIN AND BETWEEN GROUPS (ACROSS ALL PERIODS).

		Experimental condition					
		No Conflict		Low Conflict		High Conflict	
	Group	M	SD	M	SD	M	SD
Proposals [a]							
to in-group	Majority	12.87	2.47	9.82	1.95	6.50	1.27
	Minority	n/a	n/a	1.96	0.24	5.97	0.17
to out-group	Majority	n/a	n/a	0.21	0.59	0.28	0.92
	Minority	n/a	n/a	1.22	2.67	0.11	0.48
Connections [a]							
to in-group	Majority	11.95	2.90	8.79	2.17	6.06	1.41
	Minority	n/a	n/a	1.92	0.31	5.95	0.25
to out-group	Majority	n/a	n/a	0.04	0.21	0.00	0.00
	Minority	n/a	n/a	0.14	0.46	0.00	0.00
In-group favoritism [b]							
to in-group	Majority	1.00	0.00	0.98	0.07	0.95	0.15
	Minority	n/a	n/a	0.83	0.30	0.99	0.06
Reciprocation [b]							
to in-group	Majority	0.92	0.12	0.87	0.12	0.89	0.18
	Minority	n/a	n/a	0.84	0.27	0.98	0.07
Pairwise stability [b]							
to in-group	Majority	0.85	0.21	0.80	0.20	0.87	0.20
	Minority	n/a	n/a	0.96	0.15	0.99	0.04

Note:[a] Means and standard deviations for proposals and connections represent absolute numbers.

[b] Means and standard deviations for in-group favoritism, reciprocation and pairwise stability represent relative shares (percentages).

Table 4—Mixed-effects regression models on favoritism, reciprocation and connectivity

	Model A In-group favoritism		Model B Reciprocation		Model C Pairwise stability	
	B	SE	B	SE	B	SE
No Conflict (ref.)						
Low Conflict	-0.04*	(0.02)	−0.06*	(0.02)	-0.05	(0.03)
High Conflict	-0.01	(0.02)	−0.01	(0.02)	-0.01	(0.03)
Period	0.01***	(0.00)	0.03***	(0.00)	0.04***	(0.00)
Period squared	-0.000***	(0.00)	−0.001***	(0.00)	−0.001***	(0.00)
Minority (ref.)						
Majority	0.04	(0.02)	−0.04*	(0.02)	−0.14***	(0.02)
Constant	0.90***	(0.03)	0.79***	(0.02)	0.76***	(0.03)
$N_{observations}$	2,399		2,399		2,399	
$N_{individuals}$	120		120		120	
$Var_{observations}$	0.01	0.00	0.01	0.00	0.01	0.00
$Var_{individuals}$	0.01	0.00	0.01	0.00	0.01	0.00
Log likelihood	2,361.30		1,783.91		1,489.50	

Note:Unstandardized coefficients. Standard errors in parentheses. $^{*}p < 0.05,^{**}p < 0.01,^{***}p < 0.001$.

their identities the higher the level of conflict *Hypothesis 2a (Identity-dominant behavior)* they were more likely to behave as prescribed for their social category (i.e., according to their individual preference). Alternatively *Hypothesis 2b (Payoff-dominant behavior)* expected subjects to be more influenced by their social context choosing the behavior prescribed for the majority. As mentioned above, 99.3 percent of the choices corresponded to the behavior prescribed for each subject's individual preference, so that there is essentially no variation between the choices across time, subjects identity or experimental condition. Thus, the evidence suggests that regardless of the level of conflict, subjects' behavior is influenced by identities above social pressure. To summarize the findings regarding affiliation criteria and behavior adoption we present the following result:

Result 1. *In the presence of conflicting preferences, individual identities are more salient than social influence. Therefore, segregation arises between social categories.*

Hypothesis 3 (Subgame perfection) expected learning and thus increases of reciprocation with higher number of one-shot interactions, in the following referred to as period. In support of this, the positive and significant parameter estimate for period in Model B shows that reciprocation increased by 0.03 percent points with every additional interaction. This effect summed up to a total gain in 60 percent points over the whole experiment of 20 rounds. That is, identity is more salient than social influence as a behavioral criterion. By pursuing the prescribed behavior for their social category, subjects segregate. In consequence, the conflicting aspect of the interaction is put aside. The two components in the network appear as if they were two isolated populations. Once this takes place and subjects end up in a network such that those around them share their same identity (in-group), then social influence takes a relevant role again. Subjects start behaving more and more in accordance with the predictions of social influence aiming to connect with all those around them. The effect that experience and learning brings is that subjects end up decreasing the gap between the connections they propose and the connections they form, maximizing the complementarities of coordinating with their neighbors.

Similarly to the latter hypothesis, *Hypothesis 4 (Pairwise stability)* stated an increase in connections within groups with an increasing number of one-shot interactions. That is, not only subjects will coordinate more along time so that the links proposed are formed. But also, subjects will tend to form more links along time. Also supporting this assumption, the positive and significant parameter estimate for period in Model C shows that pairwise stability increased by 0.04 percent points with every additional interaction. It was reasonable to assume that the learning curve for reciprocation and pairwise stability increased steeply at the beginning and flattened out toward very high numbers of interactions, e.g. because a near-maximum had been reached in earlier interactions. The small but significant squared effects for period show indeed that both reciprocation and pairwise stability did not increase significantly anymore in later experiment periods, namely after period 10, suggesting a curvilinear learning effect. These findings are illustrated in Figure 8 and summarized in the next result:

Result 2. *In the presence of conflicting preferences, when segregation arises between social categories, subjects aim to maximize the benefits of social influence from those around them through denser networks.*

Additional tests showed that there was a learning effect in all experiment conditions, further supporting our assumptions. The learning curve was steepest in the No Conflict condition, but flattest in the High Conflict condition. The predictive margins for the different conditions are plotted in Figure 9.

Besides differences between experimental conditions and learning over periods, the regression models yielded interesting findings with regard to group membership. As presented by the negative and significant parameter estimate in Model C, subjects in the majority group reached less pairwise stability than those in the minority group. This effect occurred net of the different experimental conditions. Figure 10 shows that the difference between majority and minority group persisted throughout the entire period of the experiment. However, differences became smaller toward high numbers of one-shot interactions, which was mainly due to the learning effect in the majority group. While on average subjects in the minority reached maximum pairwise stability of 1 in period 4, subjects in the majority reached their maximum of 0.95 only in period 20. This last finding is presented in the next result:

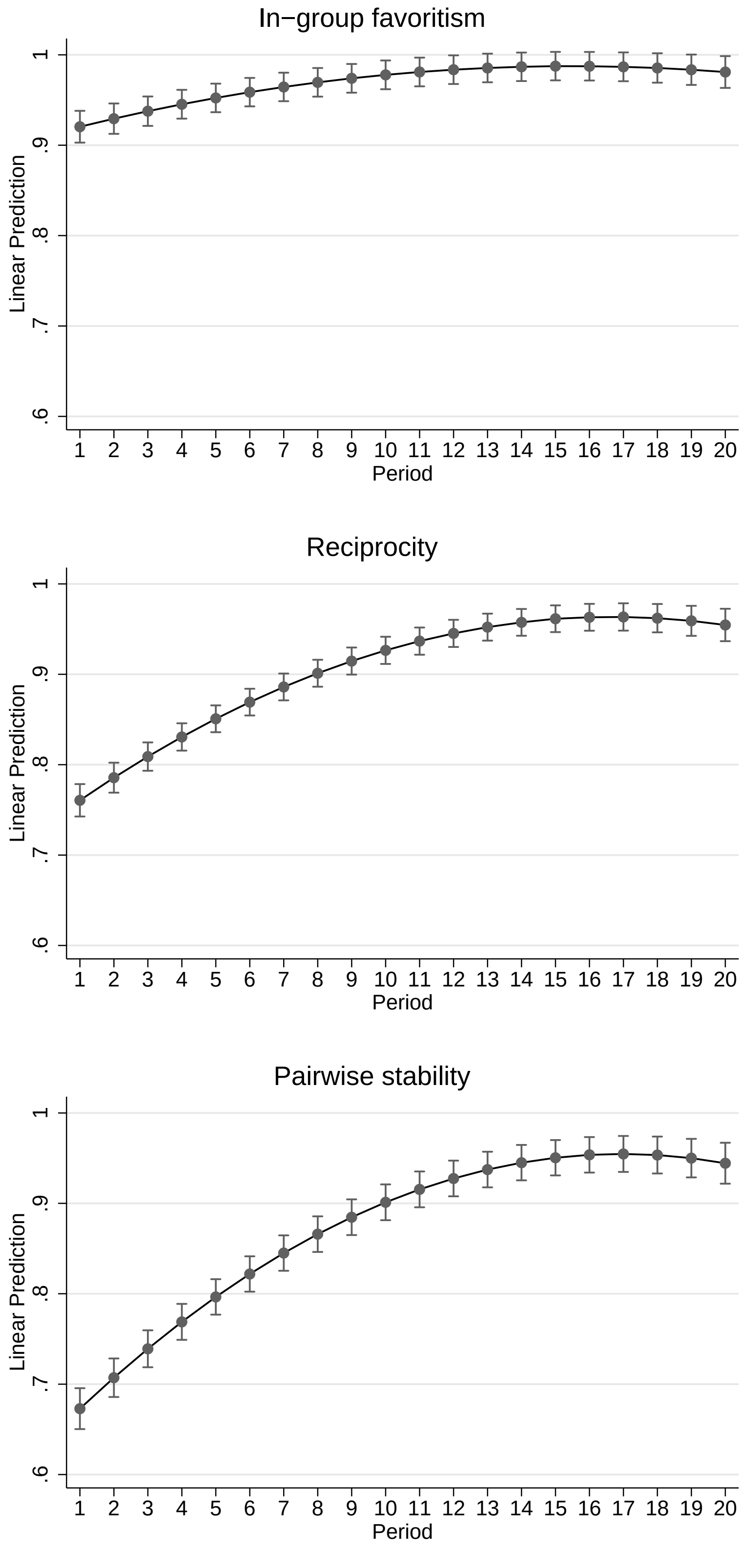


Figure 8. Predictive margins by period.

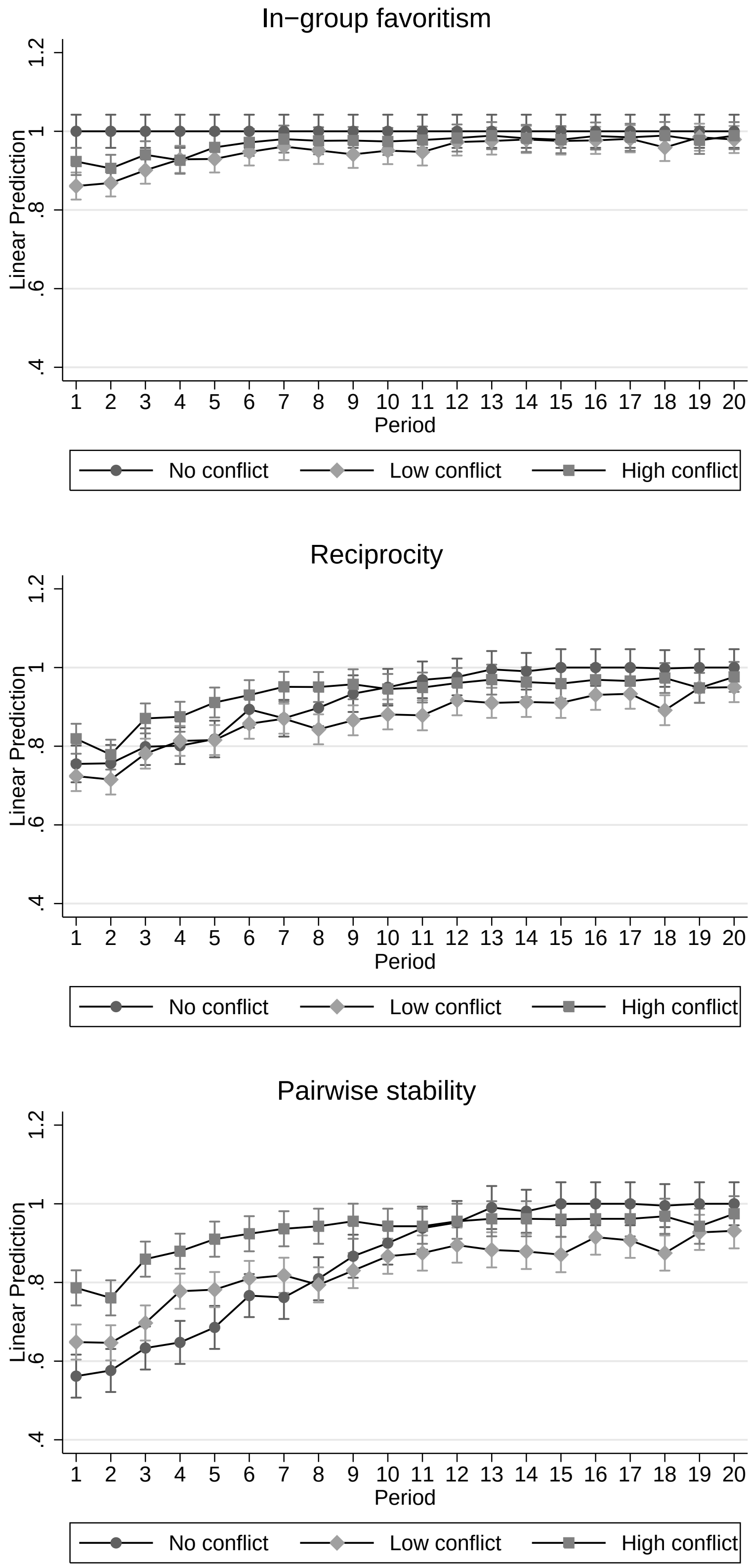


Figure 9. Predictive margins by experiment condition and period.

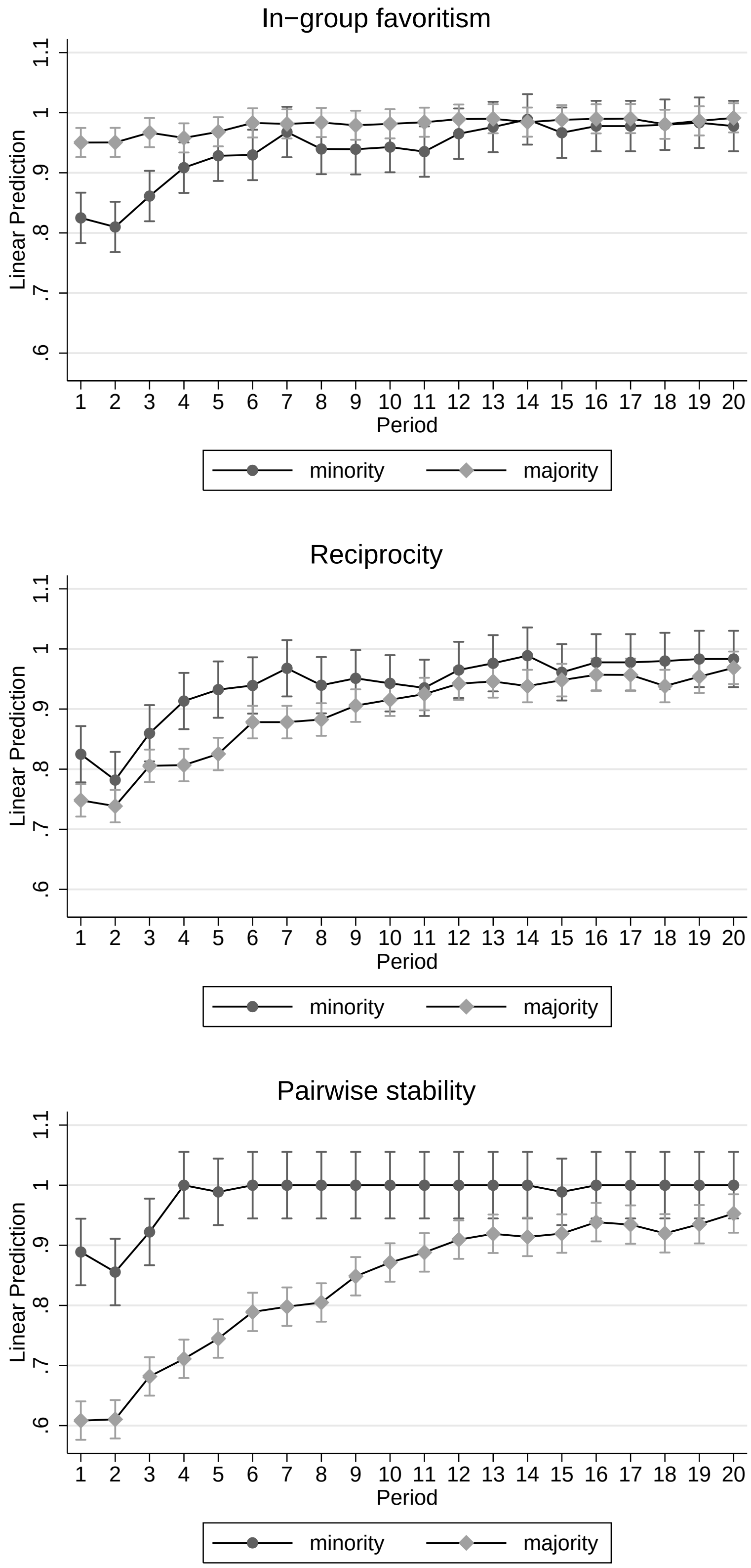


Figure 10. Predictive margins by group and period.

Result 3. *In the presence of conflicting preferences, when segregation arises between social categories, being the minority facilitates coordination and stability. Majority groups find it harder to reach affiliation consensus, which is not the case when all subjects have aligned preferences in the population.*

VI. Discussion

In this article we have argued that in social interactions with strategic complementarities the interplay between individual preferences and social influence can decisively affect outcomes of how people relate to each other and what choices they make in terms of their behavior. To elaborate this argument, we proposed a model in which actors have conflicting preferences about the behavior they want to choose but are interested in coordinating in the same behavior with more than less of those around them. Following research on identity theory (Akerlof and Kranton 2010) we characterized identities as the result of belonging to social categories. Each social category has a prescribed behavior which represent an actor's individual preference. Following research on social influence in network interactions (Jackson 2009, Hernández, Muñoz-Herrera and Sánchez 2013) we characterized in which way the behavior of others influences our decisions. To analyze the interaction of these forces that affect our decision-making process we developed a game theoretic model of network interactions with heterogeneous populations (i.e., people belonging to different social categories) and empirically tested the predictions of our theory by means of an experimental design that allowed us to control the social context, and thus the interplay between individual preferences and social influence. We first comment on our theoretical results from the model and then on the empirical findings from our experiment.

Our model, which is an extension of the work by (Hernández, Muñoz-Herrera and Sánchez 2013), indicated that the choices an actor makes about what behavior to adopt depend on her identity and the influence of others around her. An actor wants to coordinate with the highest number of neighbors making the same choice and prefers coordination on the action prescribed for her identity. As a consequence, the level of social influence needed to choose what we like is necessarily lower than the pressure we need from those around us to behave in a different way. However, this result allows for multiple outcomes depending on where in the network the influence in exerted. It is possible that all actors behave in the same way, so that the network is specialized, or that actors behave in different ways, so that the network is hybrid. Moreover, it is also possible that actors in specialized networks are all from the same social category and are all choosing what they like, so that the network is satisfactory, or there are actors from both social categories, so that some of them are not following the prescribed behavior for their identity and are frustrated.

Given the multiplicity of equilibrium outcomes that arise from games with strategic complementarities we characterized different equilibrium selection criteria. On one hand we used an adapted version of pairwise stability (Jackson and

Wolinsky 1996). If actors can coordinate bilaterally the creation of a new link, in case they are not connected, or if individually they can eliminate any relationship that is not beneficial, then only very particular network structures can result. Pairwise stable configurations would be those in which every actor is connected to all other actors who are choosing the same as her, and every actor is connected only to those choosing the same as her. This means that at the network level, the only pairwise stable configurations are either a completely connected network where every actor is behaving in the same way, or a network separated into two completely intra-connected networks, where actors in each component behave the same but not between components. Finally, we ranked the social efficiency of the resulting networks and found that the network where all actors are connected and their behavior is the same is the one that gives the highest social benefit, as long as the behavior chosen is that of the majority. This, regardless of the social composition of the population.

Using our example of the acquisition of technologies, what our model shows is that people who prefer one technology over the other will only choose what they dislike if the social pressure from those around them is stronger than the support they get to choose what is prescribed for their identity. People who like MacOS need much more pressure to buy Windows than to buy MacOS. This result is very important for threshold models because it points to the way identities and social influence interact in the adoption of different behaviors. Our model then follows to show that if people can also select the relationships they want to maintain or eliminate, the resulting network configurations are those in which MacOS users have no relation with Windows users and vice versa. Moreover, it will be likely that if actors can coordinate by pairs on what relationships to maintain, all MacOS users are linked together and all Windows users are linked together. This because individually each actor can draw out from her interactions the strategic complementarities of relating with others whose choices are compatible to theirs. As a consequence of this, the most beneficial outcome in social terms is when all actors are achieving such complementarities from all others in the population (i.e., the network is completely connected), they are all coordinating in the same behavior (i.e., the network is specialized), and the behavior is the one preferred by the majority. Thus, if MacOS users are a majority, society is at its best when all users are MacOS users even if some of them prefer Windows.

To test our theory we designed an experimental study in which we varied the composition of the population for three conditions: No Conflict, Low Conflict and High Conflict. In this way, we could assess what role individual identities and social influence play when they are interacting together but their intensity is varied. Our main empirical findings suggest that when there are different social categories, so that there are conflicting preferences about what behavior to adopt, individual identities are more salient than social influence. Therefore, networks segregate into two components. Each component has the characteristics of a satisfactory specialized network with a homogeneous population. That is, all subjects in a component belong to the same identity, they choose the behavior they prefer given their identity, and only connect with others who belong to their social category. This first result reinforces the categorization argument of identity theory showing how identities can be so strong that are used to help focalize equilibrium selection. However, the strength of individual preferences leads to two undesirable situations. In terms of relational structures, segregation between social categories is dominant. In terms of social outcomes, inefficiency is pervasive. Thus, the same force that helps individuals reduce risk and relate to others hurts society in an important way.

As a consequence, the outcome dominant in payoffs, the one that is most efficient from the societal perspective is not achievable. However, this is not because individuals are not aware of the complementarities that they could exert from relating to more than to less neighbors but because of the presence of conflict in preferences. Moreover, our second empirical result states that when there is conflict in individual preferences and segregation arises between social categories, subjects aim to maximize the complementarities from the social interactions with those around them and try to increase their number of relationships to a maximum. Thus, when a conflict between identities and social influence in relation to payoffs is latent, an actor's identity is more salient. However, once segregation emerges, so that identities are not in conflict anymore, actors social influence to each other becomes more salient so that they aim to connect completely within their component. The conflict in preferences makes the payoff dominant structure unreachable but within the segregated configuration leads to the payoff dominant case for such types of networks. This points to the tension between stability and efficiency that has been so relevant and pervasive in network studies (Jackson and Wolinsky 1996, Jackson 2009), but introduces the effect of identities in it, bowing that the stable networks emerge because of the interplay between identities and "selective" social influence. That is, only influence from those around me who are like me (in-group).

Our third empirical result is a surprising observation. When there are conflicting preferences and individuals segregate favoring only their in-groups, being in the minority facilitates coordination and stability. So, the minority groups tended to completely connect between them from early stages but the majority failed to do so until the very end of the interactions. Although this could be considered as a consequence of group size, because majorities are larger than minorities the coordination problem is greater, the failing of coordination was not present when the majority was absolute. That is, in the case of No Conflict, when all subjects belong to the majority, and grow size was the largest, they did not show the same limitations in maximizing the complementarities of their social connections by reaching pairwise stable networks. This result complements the existing work on in-group bias in identity theory.

As mentioned before, when identification is experimentally induced (i.e., minimal group paradigm), in-group bias has been significantly observed. (Leonardelli and Brewer 2001) even observed this for cases where there was a majority and a minority. Our results complement these findings on the literature by showing that in network interactions with conflicting preferences in-group bias is observed but groups in numerical minorities express more bias than those in numerical majorities. Our results go in accordance with what has been empirically found by (Mullen, Brown and Smith 1992, Otten, Mummendey and Blanz 1996) who have observed, outside of the lab, bias in group size both when groups are real or artificial.

Some potential limitations of our work warrant further discussion. Compared to other works on identities (Akerlof and Kranton 2000), we model social categories as fixed while their works have assumed that individuals can choose their individual identity and not only their behavior. Our main aim was to understand the adoption of behavior when given identities and social influence are at play in context of conflicting preferences. Accordingly, we decided to maintain the identity assumptions central to our approach. Fixed social categories are common in research on identities (i.e., race, gender, nationality) and our model can be extended to include variable identities in further research.

REFERENCES

Abrams, Dominic, and Michael A Hogg. 2012. *Social identifications: A social psychology of intergroup relations and group processes.* Routledge.

Akerlof, George A, and Rachel E Kranton. 2000. "Economics and identity." *The Quarterly Journal of Economics*, 115(3): 715–753.

Akerlof, George A, and Rachel E Kranton. 2002. "Identity and schooling: Some lessons for the economics of education." *Journal of economic literature*, 40(4): 1167–1201.

Akerlof, George A, and Rachel E Kranton. 2005. "Identity and the Economics of Organizations." *The Journal of Economic Perspectives*, 19(1): 9–32.

Akerlof, George A, and Rachel E Kranton. 2010. *Identity economics: How our identities shape our work, wages, and well-being.* Princeton University Press.

Bala, V., and S. Goyal. 2000. "A noncooperative model of network formation." *Econometrica*, 68(5): 1181–1229.

Ballester, Coralio, Antoni Calvó-Armengol, and Yves Zenou. 2006. "Who's who in networks. wanted: the key player." *Econometrica*, 74(5): 1403–1417.

Bernhard, Helen, Ernst Fehr, and Urs Fischbacher. 2006. "Group affiliation and altruistic norm enforcement." *The American Economic Review*, 217–221.

Billig, Michael, and Henri Tajfel. 1973. "Social categorization and similarity in intergroup behaviour." *European Journal of Social Psychology*, 3(1): 27–52.

Brewer, Marilynn B. 1979. "In-group bias in the minimal intergroup situation: A cognitive-motivational analysis." *Psychological bulletin*, 86(2): 307.

Calvó-Armengol, Antoni, and Rahmi İlkılıç. 2009. "Pairwise-stability and Nash equilibria in network formation." *International Journal of Game Theory*, 38(1): 51–79.

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

Charness, Gary, Luca Rigotti, and Aldo Rustichini. 2007. "Individual behavior and group membership." *The American Economic Review*, 1340–1352.

Chen, Yan, and Sherry Xin Li. 2009. "Group identity and social preferences." *The American Economic Review*, 431–457.

Eckel, Catherine C, and Philip J Grossman. 2005. "Managing diversity by creating team identity." *Journal of Economic Behavior & Organization*, 58(3): 371–392.

Ellison, Glenn. 1993. "Learning, local interaction, and coordination." *Econometrica*, 61: 1047–1047.

Fischbacher, Urs. 2007. "z-Tree: Zurich toolbox for ready-made economic experiments." *Experimental economics*, 10(2): 171–178.

Galeotti, A., S. Goyal, M.O. Jackson, F. Vega-Redondo, and L. Yariv. 2010. "Network games." *Review of Economic Studies*, 77(1): 218–244.

Goette, Lorenz, David Huffman, and Stephan Meier. 2006. "The impact of group membership on cooperation and norm enforcement: Evidence using random assignment to real social groups." *The American economic review*, 212–216.

Goyal, Sanjeev. 2007. *Connections: an introduction to the economics of networks.* Princeton University Press.

Granovetter, Mark. 1978. "Threshold models of collective behavior." *American journal of sociology*, 83(6): 1420.

Hernández, Penélope, Manuel Muñoz-Herrera, and Ángel Sánchez. 2013. "Heterogeneous network games." *Games and Economic Behavior*, 79: 56.

Jackson, Matthew O. 2009. "Networks and economic behavior." *Annu. Rev. Econ.*, 1(1): 489–511.

Jackson, Matthew O, and Alison Watts. 2001. "The existence of pairwise stable networks." *Seoul Journal of Economics.*

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

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

Kandori, Michihiro, George J Mailath, and Rafael Rob. 1993. "Learning, mutation, and long run equilibria in games." *Econometrica: Journal of the Econometric Society*, 29–56.

Lazarsfeld, Paul F, Robert K Merton, et al. 1954. "Friendship as a social process: A substantive and methodological analysis." *Freedom and control in modern society*, 18(1): 18–66.

Leonardelli, Geoffrey J, and Marilynn B Brewer. 2001. "Minority and majority discrimination: When and why." *Journal of Experimental Social Psychology*, 37(6): 468–485.

López-Pintado, Dunia. 2006. "Contagion and coordination in random networks." *International Journal of Game Theory*, 34(3): 371–381.

Marsden, Peter V. 1990. "Network diversity, substructures, and opportunities for contact." *Structures of power and constraint: Papers in honor of Peter Blau*, 397–410.

McLeish, Kendra N, and Robert J Oxoby. 2007. "Identity, cooperation, and punishment." IZA Discussion Papers.

McPherson, Miller, Lynn Smith-Lovin, and James M Cook. 2001. "Birds of a feather: Homophily in social networks." *Annual review of sociology*, 415–444.

Morris, Stephen. 2000. "Contagion." *The Review of Economic Studies*, 67(1): 57–78.

Mullen, Brian, Rupert Brown, and Colleen Smith. 1992. "Ingroup bias as a function of salience, relevance, and status: An integration." *European Journal of Social Psychology*, 22(2): 103–122.

Muñoz-Herrera, Manuel, Jacob Dijkstra, Andreas Flache, and Rafael Wittek. 2013. "How specialization can breed social exclusion: A model of strategic interaction between pecialists and generalists in knowledge-intensive productive exchange." *Working Paper, University of Groningen.*

Otten, Sabine, Amelie Mummendey, and Mathias Blanz. 1996. "Intergroup discrimination in positive and negative outcome allocations: Impact of stimulus valence, relative group status, and relative group size." *Personality and Social Psychology Bulletin*, 22(6): 568–581.

Schelling, Thomas. 1978. *Micromotives and Macrobehavior.* New York: WW Norton.

Snijders, T, and Roel J Bosker. 2012. *Multilevel analysis: An introduction to basic and applied multilevel analysis.* . 2nd ed., London: Sage.

Tajfel, Henri, and John C Turner. 1979. "An integrative theory of intergroup conflict." *The social psychology of intergroup relations*, 33: 47.

Tajfel, Henri Ed. 1978. *Differentiation between social groups: Studies in the social psychology of intergroup relations.* Academic Press.

Tanaka, Tomomi, and Colin F Camerer. 2009. "Status and Ethnicity in Vietnam: Evidence from Experimental Games." In *Social Computing and Behavioral Modeling.* 1–2. Springer.

Turner, John C. 1978. "Social categorization and social discrimination in the minimal group paradigm." *Differentiation between social groups: Studies in the social psychology of intergroup relations*, 101–140.

Turner, John C, Michael A Hogg, Penelope J Oakes, Stephen D Reicher, and Margaret S Wetherell. 1987. *Rediscovering the social group: A self-categorization theory.* Basil Blackwell.

Vega-Redondo, F. 2007. *Complex social networks.* Vol. 44, Cambridge Univ Pr.

Verbrugge, Lois M. 1977. "The structure of adult friendship choices." *Social Forces*, 56(2): 576–597.

Vives, Xavier. 1990. "Nash equilibrium with strategic complementarities." *Journal of Mathematical Economics*, 19(3): 305–321.

Vives, Xavier. 2005. "Games with strategic complementarities: New applications to industrial organization." *International Journal of Industrial Organization*, 23(7): 625–637.

Young, H Peyton. 1993. "The evolution of conventions." *Econometrica: Journal of the Econometric Society*, 57–84.

Zhao, Jijun, Miklos N Szilagyi, and Ferenc Szidarovszky. 2008. "An n-person battle of sexes game." *Physica A: Statistical Mechanics and its Applications*, 387(14): 3669–3677.

Appendix: Subjects' instructions (For online publication)

INSTRUCTIONS

Welcome. You are going to participate in an economic experiment. Please read carefully the following instructions. If you have any question, please rise up your hand and one of the experimenters will answer your question personally. During the experiment you are not allowed to communicate with other participants, neither to use your cellphone, nor to use the computer for anything else but to participate in the experiment.

In this experiment you will earn points, which will exchanged into euros. The number of points you earn depend on your choices and the choices of the other participants.

You will participate in this experiment for 25 rounds. The first 5 rounds are trial rounds. At the beginning of the experiment all the participants are randomly divided into groups of 15 people, each person identified with a number from 1 to 15 (that is 1,2,3,…,15). The computer will randomly assign a symbol to each participant: *circle* or *square*. The symbol for each participant will be kept constant along the entire experiment, but the group composition and the numbers will change randomly in each round. All participants will be informed about the number and symbol assigned to each and every member in their group, but they will not be informed of their identity.

In each round you will make 2 decisions, and every round consists of 4 parts:

1. You choose, from the 14 remaining members in your group, whom you want to propose a connection to.
2. You will be informed of the connection proposals all members in your group make to you and to others. A connection is formed between two participants if both have proposed a connection to each other.
3. You will choose an action: *up* or *down*.
4. You will be informed of action chosen by all members of your group and of the points you have earned in that round.

Your decisions, connection proposal and action, as well as the decisions of the other participants in your group, will determine the total number of points you can earn in each round. **Each connection proposal, even if it is not formed, costs 2 points**. You will receive points for each participant you are connected to who chooses the same action you choose. The number of points you earn depends on the action that is chosen and on your symbol:

You are ***circle***:

- If you choose ***up*** you get **6 points for each** coordination with your connections.
- If you choose ***down*** you get **4 points for each** coordination with your connections.

You are ***square***:

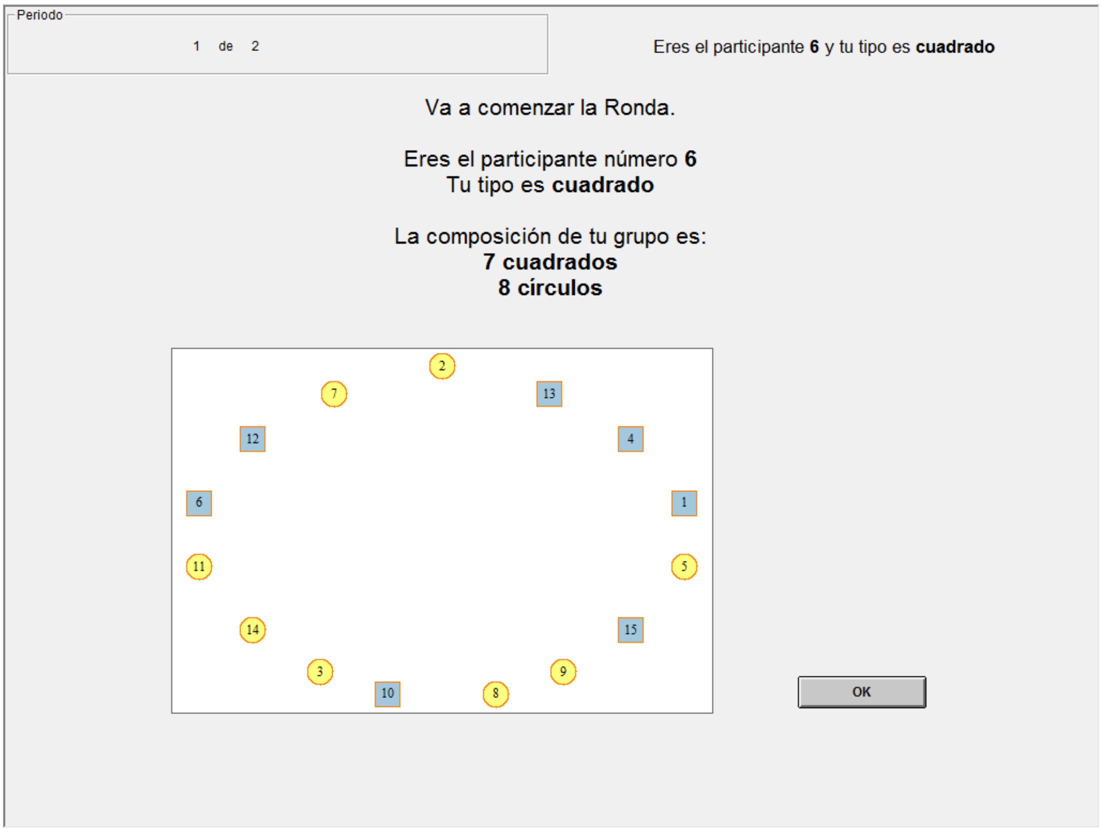
- If you choose ***down*** you get **6 points for each** coordination with your connections.
- If you choose ***up*** you get **4 points for each** coordination with your connections.

The detailed instructions for each part of the experiment together with some examples are presented as follows (keep in mind the numbers of each participant, their symbol, the connections formed and the actions chosen are only examples and need not occur in this same way along the experiment):

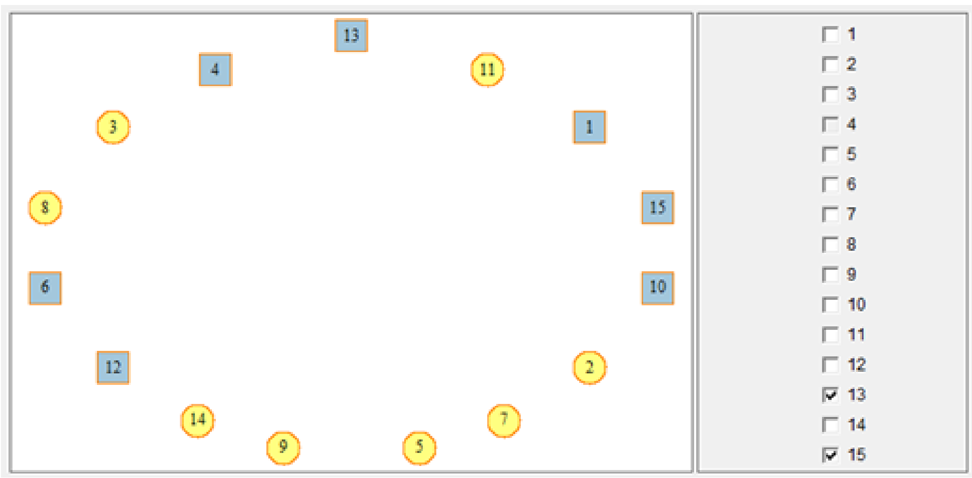
Beginning of a round:

At the beginning of each round you will be informed of your number and symbol, as well as the numbers and symbols of all other participants in your group. Your symbol will be the same along the entire experiment, but your number and the composition of your group will change.

[***Text in the image***: The round is about to start. You are participant number 6. Your symbol is **square**. The composition in your group is **7 squares 8 circles**.]

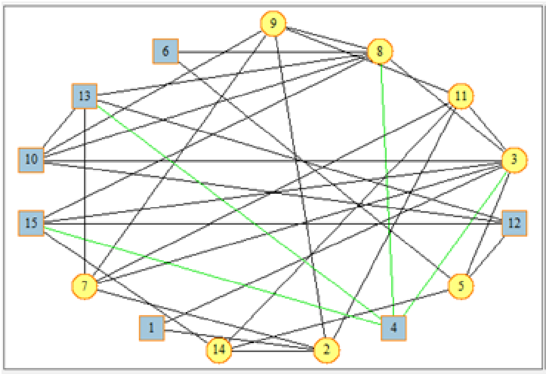


Part 1- Proposals:



The first decision you make is whom in your group you want to propose a connection to. To propose a connection you check the box next to the number of a participant on the list at the right hand side. In the example above, connections are proposed to participants 14 and 15.

Part 2- Connections:



Proposiciones Realizadas

1		2	3	4		6	7		9		11		13	14	
2	1				5		7	8	9		11		13	14	
3	1	2		4	5		7	8		10	11	12			15
4		2	3		5			8			11		13		15
5			3			6		8				12		14	
6				4	5			8		10		12	13		15
7		2	3			6		8	9	10	11		13		15
8			3	4		6	7		9	10			13		15
9		2	3	4	5		7	8		10	11		13		15
10	1		3	4				8	9			12	13		
11		2	3		5		7	8	9				13	14	15
12				4	5		7		9	10			13		15
13			3	4	5		7	8		10		12			
14		2			5	6		8	9		11				15
15		2	3	4	5			8		10		12		14	

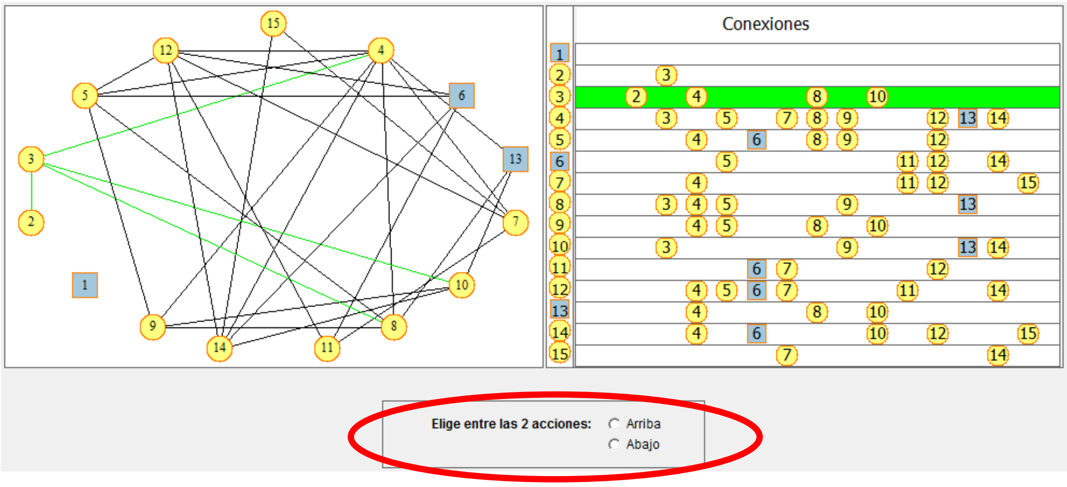
Once all the participants have proposed the connections they want to form you will see the network of connections that results. A connection is formed when two participants propose to each other. In the picture of the network you will see your connections highlighted in green, and on the table next to it you will also see a Green highlight on the row corresponding to the proposals you have made and on the column corresponding to the proposals you have received.

In the example above you are participant **4**. You have proposed connections to the following participants:

- Participants with symbol circle: 2,**3**,5,**8**, and 11
- Participants with symbol square: **13** and **15**

You have been proposed a connection from participants 1, **3**, 6, **8**, 9, 10, 12, **13** and **15**. Therefore, you have a connection with 3, 8, 13 and 15, which are the participants to whom you proposed a connection and who also proposed a connection to you. That is, the final connections are the intersection between the proposals made and received.

Part 3- Action:



[***Text in the image***: Choose between the 2 actions: *Up* - *Down*]

Once the network of connections is formed, in the next part you will choose an action, *up* or *down*. You will be able to see the formed network, but on the table on the right hand side you will only see the connections formed between the participants. Remember that the points you can earn are:

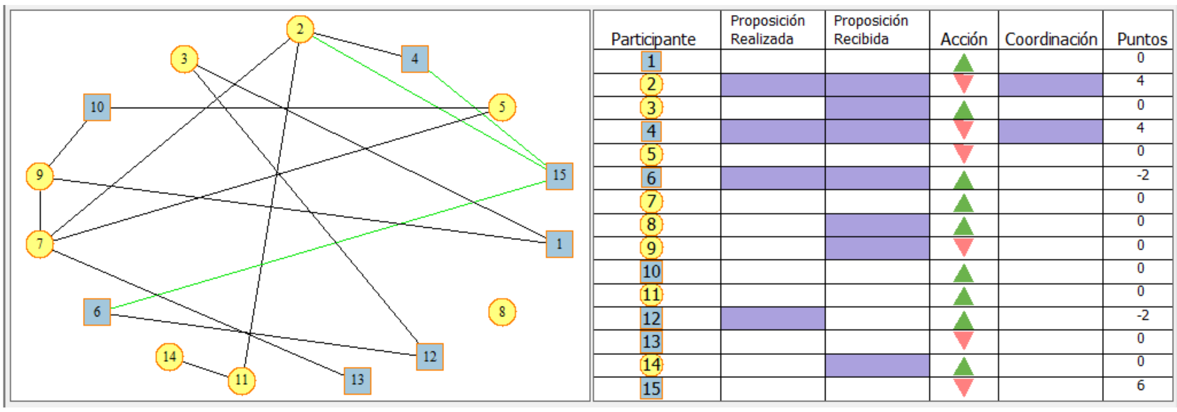
If you are a **circle**:

- 6 points for each coordination in ↑
- 4 points for each coordination in ↓

If you are a **square**:

- 6 points for each coordination in ↓
- 4 points for each coordination in ↑

Part 4- Summary:



In the last part you will see a screen summarizing what happened in the current round: the proposals you made, the proposals made to you, the action chosen by every participant, whether you coordinated or not, and the points you earned from the interaction with each one of them.

In the example above you are participant **15**, your symbol is *square*, and you have chosen action *down*. Let's observe in detail you interaction with some of the participants:

- With **2**: You proposed and connection and 2 proposed a connection to you, and you have coordinated in the action you chose. You win: 6 – (cost of the proposal) = 6-2=4 points
- With **6**: There is a connection formed by you did not coordinated. You earn no points but pay the cost of the proposal (-2 points)
- With **12**: You proposed a connection but 12 did not propose to you. You pay the cost of the proposal (-2 points)
- With **14**: You did not propose a connection but 14 proposed a connection to you. 14 pays the cost of proposing a connection but you do not pay.
- With **15**: You always coordinate with "yourself"

Below you can see a table that can help you calculate the total number of points you can earn by coordinating with those you are connected to (the cost of the proposals are **NOT** subtracted, remember that each proposal costs **2** points):

Si eres ○ y elijes △
Si eres □ y elijes ▽

Coordinaciones	Puntos
0	6
1	12
2	18
3	24
4	30
5	36
6	42
7	48
8	54
9	60
10	66
11	72
12	78
13	84
14	90

Si eres ○ y elijes ▽
Si eres □ y elijes △

Coordinaciones	Puntos
0	4
1	8
2	12
3	16
4	20
5	24
6	28
7	32
8	36
9	40
10	44
11	48
12	52
13	56
14	60

[***Text in the LEFT column***: If you are CIRCLE and choose UP – If you are SQUARE and choose DOWN – Coordinations – Points
Text in the RIGHT column: If you are CIRCLE and choose DOWN – If you are SQUARE and choose UP – Coordinations - Points]