The Impact of individual differences on network relations
Muñoz Herrera, Manuel

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2015

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
Conflict and segregation in networks: An experiment on the interplay between individual preferences and social influence

5.1 Introduction

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. For instance, when choosing our friends (McPherson et al., 2001) 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). Examples of social influence range from which products we buy (Galeotti et al., 2010), whether we engage or not in criminal activities (Ballester et al., 2006), to our participation in collective action (Granovetter, 1978). Our aim in this chapter is to understand the forces motivating how people decide what relationships to form and how to behave with others by studying the interplay between individual preferences and social influence.

One of the most prominent theoretical tools to study the effect individual preferences have on the way people behave is identity theory (Akerlof and Kranton, 2000; Tajfel and Turner, 1979). 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

\[\text{identity} = \text{categorization} \times \text{identification}\]
we identify with is the *in-group*. Conversely, the group we do not identify with is the *out-group*. Third, *comparison*, the process we use to compare our *in-group* and the *out-group*. The social categories people identify with are associated with a particular order in how actors rank the available choices. We refer to this order in ranking choices as a person’s individual preference. When people are doing what is in accordance to their individual preferences they get more out of it, and those who are not choosing what they prefer, given 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 we make our choices influenced by our social relationships is that of *strategic interaction in networks* (i.e. coordination games in networks). For instance, if a person is choosing a technological product and wants it to be compatible with her co-workers, her choice can change depending on how many of them are using the same technology or a different one (Vives, 2005). These interactions are known as coordination games with strategic complementarities, where a person’s incentives to 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, people are more likely to adopt a given behavior or not depending on how others behave, even if such a behavior is not the one they rank highest given their identity (Hernández et al., 2013).

The existing research on identity theory and coordination network games, 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 chapter aims to address this gap and give account of the interplay between individual preferences and social relationships. A key aspect of the relationships we model is that they portray strategic complementarities. Actors are better off aligning their behavior to that of those around them. However, their identities introduce a conflict about which behavior each prefers to adopt. Thus, we model the interplay between identities and social influence in a context of *conflicting preferences*. 

To do so, we elaborate and analyze a formal model in which actors choose with whom to interact and which behavior to adopt. Our model moves beyond the existing work in its combination of three features. First, it introduces identities as part of the strategic considerations actors have. Second, to assess the effect of identities on the establishment of relationships, actors choose their social network. Third, to understand how social influence affects actors choices, the adoption of behavior is made once the structure of relationships has been formed. There is one behavior ranked highest to each social category, so that the preference of an individual is to adopt the behavior that gives her most benefit in relation to her identity, but there is a benefit in behaving the way those around her 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.
We also design and run an experiment derived from our theoretic model, in which subjects are artificially assigned an identity, and the composition of identities in the group is known to everyone. We vary the relative size of the social categories, which changes the intensity of the conflict in preferences between subjects. Experimental conditions range from No Conflict, when every subject has the same identity, through Low Conflict, when there is a large majority and a small minority, to High Conflict, when the size of the two groups is almost the same. By means of these variations we can study how differences in the level of conflict between the subjects participating (i.e. the relative size of the groups with one preference or another) can lead to social exclusion (i.e. segregation) and inequality between them.

The remainder of this chapter builds as follows: In Section 5.2 we describe our theoretical framework. The game theoretic model is presented in Section 5.3. Section 5.4 analyzes the network structures that emerge from the interactions of actors belonging to different social categories. In Section 5.5 we describe the experimental study. Section 5.6 presents the main results of our experiment. We conclude with a discussion of the implications and limitations of the study in Section 5.7.

## 5.2 Framework

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, (Akerlof and Kranton, 2010) and strategic interaction in networks (from economics). Our study integrates both lines of research for interactions with strategic complementarities.

### 5.2.1 Identities and social influence

Research on the theory of identities was initiated in psychology (Tajfel and Turner, 1979; Turner, 1987), mainly focusing on the effects that the social context has on group processes and inter-group relations. The aim being to understand how inter-group interactions could be explained and whether groups of people who share/differ in certain traits were more likely to integrate/discriminate between them. 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, so they think of their own group as good.

The argument of in-group bias has been widely supported by experimental research on identities (Billig and Tajfel, 1973; Ellemers et al., 1999). To assess the effect of identities on inter-group relations, the methodology commonly used is the minimal group paradigm, which seeks for minimal conditions that would create group identification. Subjects are assigned to groups using arbitrary criteria (i.e. preference between paintings). After
informing subjects of their group membership (i.e. their identity), they were asked to allocate points to members of their in-group and to members of their out-group. Minimal group experiments have typically shown a tendency to allocate more points to in-group than to out-group members (Brewer, 1979; Mullen et al., 1992), illustrating the strong tendencies that group identification generate on our individual preferences. An important limitation is that this approach has no strategic considerations about the way people behave given the behavior of others. 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). The interaction of identity considerations and individual incentives had not been directly addressed theoretically or experimentally, in psychology, leaving an important gap to be developed.

George 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). The application of their model has been found useful to explain gender discrimination (Akerlof and Kranton, 2000), why costs and benefits of education are not enough to explain who is enrolled in school and their choices of working hard at it (Akerlof and Kranton, 2002), and why a firm operates well when employees identify with it and why monetary incentives alone do not work (Akerlof and Kranton, 2005). A set of experimental work has also included identity as part of the analysis, addressing the limitation in the psychological approach by taking into account monetary stakes (Bernhard et al., 2006; Charness et al., 2007; Chen and Li, 2009; Goette et al., 2006). 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. 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 understanding the way others influence our behavior.

Research on networks introduced the strategic behavior of people into the analysis of social influence by modeling interactions as games (for surveys of the literature see Goyal 2007; Jackson 2010; Vega-Redondo 2007). Network games model the way individuals behave as a function of the actions of their neighbors. In settings where individuals are better off the more of their neighbors behave as they do but there are at least two possible behaviors, influence is captured by thresholds functions (Galeotti et al., 2010; Granovetter, 1978). 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 the same option or some acquire one technology and some acquire the other. Work following this aim are Ellison (1993); Kandori et al. (1993); López-Pintado
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. 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.

Experimental studies on coordination games in networks provide empirical evidence that the network structure affects choices. (Keser et al., 1998) compared circle structures and three-person groups and found evidence that in the latter subjects are more likely to coordinate in the risk dominant equilibrium. The effect of local interactions and the play of risk dominant or payoff dominant equilibria have been studied by (Berninghaus et al., 2002) and (Cassar, 2007). In particular, (Cassar, 2007) found that risk-dominant equilibria are more likely to be played in “small-world” network structures than in random networks or in networks with high clustering, where neighbors overlap.

Two main aspects of this research line 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 of these limitations has received a great deal of 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 Bala and Goyal (2000); Jackson and Watts (2002); Jackson and Wolinsky (1996); Muñoz Herrera et al. (2014). A main finding that endogenous formation brings to network games is that the risk-dominant equilibrium is not the most salient equilibrium anymore. If actors can choose with whom they want to affiliate, this reduces risk and the payoff dominant equilibrium becomes salient (Jackson and Watts, 2002). The idea is that people act strategically when deciding with whom to form social relationships. Thus, the strength of social influence can vary depending on whether we are able to adapt our relationships with others given what we are interested in choosing.

The second of these limitations, the inclusion of identities in network settings, has not received much attention. A study of conflicting preferences, closely linked to ours, is the work by (Hernández et al., 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, so that actors have no choice regarding whom they relate to. We extend their work 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 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 people’s behavior, given the interplay between individual preferences
and social influence.

Two papers our work is closely related to are Bojanowski and Buskens (2011) and Corten and Buskens (2010). Bojanowski and Buskens (2011) investigates the strategic interaction of heterogeneous actors in coordination environments (i.e. networks). This work presents a very interesting characterization of stability in which no actor wants to change her relationships or her behavior (i.e. chosen convention) and no pair of actors who are not related have incentives to form a new relationship. Furthermore, apart from the analytical work, their paper provides computer simulations to study how heterogeneity affects the emerging relationships. A main difference between the theoretical model in Bojanowski and Buskens (2011) and the model we present in this study relates to the focus each study gives to the meaning of choosing according to one’s preferences. While in their work, an actor earns a payoff regardless of whether she coordinates in the behavior with a partner or not, as long as she chooses her preferred behavior (i.e. the authors refer to it as native), in this Chapter, as well as in Chapter 4, the condition is that actors only benefit from a relationship if they both achieve to establish a link between them and to coordinate in the behavior they adopt. This difference between the two studies makes the results complementary, and is due to the salience that the conflict between preferences and influence has in our study. In the Section 5.7 we comment how our results connect to those in Bojanowski and Buskens (2011).

Corten and Buskens (2010) experimentally studies how coordination networks emerge when different conventions are available to choose from. Actors start from an exogenously-given structure and are able to choose their relationships for repeated interactions. This work provides a full-fledged analysis where they approach the problem analytically, provide computer simulations to assess their model and empirically test it through an experimental study where actors can choose both their relationships and their behavior.

An essential difference between Corten and Buskens (2010) and our work in this chapter is that the actors are heterogeneous in the population. That is, in the line of Bojanowski and Buskens (2011) and the work we provide in Chapter 4, we complement Corten and Buskens (2010)’s experimental approach by studying the effect of heterogeneity in the patterns of relationships that subjects form in the experimental tasks. In the same way, our experimental work complements the computer simulations presented in Bojanowski and Buskens (2011). That is, our work on coordination networks complements and extends our understanding of how conflicting preferences affects choices on behavior and partner selection in coordination networks.

In conclusion, this Chapter proposes a model that extends the theoretical work in Chapter 4 and develops an experimental study of the findings from the model. Particularly, the extension of the theoretical model is on the way network relationships occur. In Chapter 4 the focus was on the effect of conflicting preferences in network games, so that network relationships were not part of the individual decisions of the actors but given exogenously to them. In Chapter 5 we relax this assumption and model situations in which actors can choose what behavior to adopt and can also decide who they want to interact with. Thus, in our model actors play a coordination game with multiple partners and can decide who
those partners are. We study stable outcomes in which actors can deviate unilaterally (i.e. Nash equilibria) and also in pairs (i.e. Pairwise stable Nash equilibria). In our model we investigate how exclusion of potential partners and inequalities in the benefits acquired by the actors depend on the level of heterogeneity in the network.

The second part of this chapter presents the results of a laboratory experiment that tests the analytical results of the model. Particularly, groups of 15 actors play a coordination game and choose with whom to form a relationship and what behavior to adopt from two possible choices. Actors earn higher benefits by coordinating in one action or another, depending on their preference. Based on this scheme of incentives, we design three experimental treatments that vary the level of conflict in the network due to the level of heterogeneity in the population. The treatments are labeled as no conflict if all actors have the same preference, low conflict if there is a significant majority of actors with one preference and a very small minority of actors with the opposite preference, and high conflict when majority and minority are not too different in size. Given the interplay between the incentives an actor has to choose the behavior she prefers, and the incentives an actor has to choose alike with those around her, our experiment studies under what conditions segregation between actors and inequality in their earnings result, form the level of conflict in their preferences.

5.3 The model

In this section we present our model of network interactions where 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 an undirected network, and play a network game denoted by $\Gamma$. In $\Gamma$ 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 $\Gamma$ 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 $p_i^1$ in $\{0, 1\}^N$, where $p_i^j = 1$ means that player $i$ proposes a link to $j$, and $p_i^j = 0$ otherwise. We assume $p_i^i = 0$. Only if $p_i^j = p_i^j = 1$, we say there is a link between $i$ and $j$. The profile of vectors $p = (p_1^1, p_2^2, \ldots, p^n)$ represents the network by the set of links, $g$. If a pair of players $i$ and $j$ are connected by a link, it is denoted as $g_{ij} = 1$, and if there is no link between them, we say $g_{ij} = 0$. The degree of player $i$ is represented by the set of neighbors she has, $k_i(g) = \{j : g_{ij} = 1\}$, with cardinality $k_i$. 
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 $i$, $x_i$, 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 through the linear payoff function, $u_i$, that strategically depends on the choices made by connected players (we denote $x_k(g)$ as the vector of actions taken by $i$’s neighbors), their identities and proposed links in the first stage, as follows in Equation 5.1:

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

(5.1)

where $I_{\{x_j = x_i\}}$ is the indicator function of those neighbors choosing the same action as player $i$. The parameter $\lambda$ is defined by $\lambda^{\theta_i} = \alpha$ when a player chooses the action prescribed for her identity, and $\lambda^{\theta_i} = \beta$ otherwise. The cost of proposing a link is $c > 0$, and the relation between the parameters in the model is $0 < c < \beta < \alpha$.

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, extending the applicability of network models to situations in which the preferences of different players may be in conflict.

In order to study the equilibrium of the sequential game, we fix a network configuration $\{g\}$ generated by the profile $p$. In the second stage, 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:

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

Note that $u_i(\theta_i, p, x_1^*, \ldots x_n^*) = u_i(\theta_i, p, x_1^*, x_k^*(g))$, the actions of players that are not $i$’s neighbors do not change her payoff.

### 5.4 Equilibrium

In this section we provide the Nash equilibrium characterization for our network game, $NE(\Gamma)$. To do this we follow Hernández et al. (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.$^1$

### 5.4.1 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 a player chooses the action she prefers (satisfactory) or the disliked action (frustrated).$^2$ Thus, there are four possible configurations: (i) satisfactory specialized ($SS$) where all players coordinate on the same action, which is their preferred choice; (ii) frustrated specialized ($FS$), where all players coordinate on the same action, but at least one of them is choosing her disliked option; (iii) satisfactory hybrid ($SH$), where all 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 ($FH$) which portray both actions and at least one player chooses her disliked option.

### 5.4.2 Nash equilibrium

Once the network is realized, the results in Hernández et al. (2013) for fixed networks are applicable to our case. There are two threshold functions when players have conflicting preferences. The function $\tau(k_i)$ 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 $\overline{\tau}(k_i)$ represents the maximum number of neighbors choosing the non-favorite action so that $i$’s best response is still to adopt the behavior she likes. If one more of her neighbors chooses the non-favorite action, $i$’s best response is to adopt her disliked option. Proposition 5 shows this, where the number of $i$’s neighbors choosing action 1 is $\chi_i$ and the number of her neighbors choosing action 0 is $k_i - \chi_i$.$^3$

$^1$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 zero, given the cost of proposal is independent of the action played in the second stage. 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: $[u_i(1,p',1,x_N,(g)) - cp'] - [u_i(1,p',0,x_N,(g)) - cp'] = u_i(1,1,x_N,(g)) - u_i(1,0,x_N,(g))$. Therefore, this cost is cancelled on both sides of the computation.

$^2$We denote 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 why $\alpha > \beta$.

$^3$Because the threshold functions represent a number of neighbors, we denote by $\lceil \ldots \rceil$ and $\lfloor \ldots \rfloor$ respectively the maximum integer and maximum lower integer of the real number considered.
Proposition 5 (Hernández et al., 2013) Let
\[
\tau(k_i) = \left\lceil \frac{\beta}{\alpha + \beta} k_i - \frac{\alpha - \beta}{\alpha + \beta} \right\rceil,
\]
\[
\tau(k_i) = \left\lfloor \frac{\alpha}{\alpha + \beta} k_i + \frac{\alpha - \beta}{\alpha + \beta} \right\rfloor,
\]
developed 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
\[
x^*_i = \begin{cases} 
1, & \text{iff } x_i \geq \tau(k_i), \\
0, & \text{otherwise}.
\end{cases}
\]
The best response of player \( i \) with identity \( \theta_i = 0 \) and degree \( k_i \), \( x^*_i \), is
\[
x^*_i = \begin{cases} 
0, & \text{iff } x_i \leq \tau(k_i), \\
1, & \text{otherwise}.
\end{cases}
\]
The intuition behind Proposition 5 is that 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. Clearly, a player \( i \) requires less influence from her social network to choose what she prefers and more social pressure to adopt her disliked behavior \( (\tau(k_i) > \tau(k_i)) \), compared to an analysis ignoring identities. For instance, in our 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 friends to purchase this operative system. However, they would require more pressure from their friends to buy the Windows system that they dislike.

### 5.4.3 Subgame perfect Nash equilibrium

So far we have 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 a link back to \( i \). The first Lemma states that in a SPNE for a given network \( \{g\} \) the vector of proposed links \( p^i \) for any player \( i \) does not exceed the set of her realized neighbors in \( \{g\} \).

**Lemma 3** Subgame perfect Nash equilibria: 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:
• \(p^i\) with \(p^i_j = 1\) if \(j \in \{i_1, i_2, \ldots, i_k\}\) and \(p^i_j = 0\) if \(j \notin \{i_1, i_2, \ldots, i_k\}\)

• \(\tilde{p}^i\) with \(\tilde{p}^i_j = 1\) if \(j \in \{i_1, i_2, \ldots, i_k, z_1, z_2, \ldots, z_s\}\) and \(\tilde{p}^i_j = 0\) if \(j \notin \{i_1, i_2, \ldots, i_k, z_1, z_2, \ldots, z_s\}\).

where the set of players \(\{z_1, z_2, \ldots, z_s\} \cap \{i_1, i_2, \ldots, i_k\} = \emptyset\)

For the game \(\Gamma\) where \(\{g_{ij}\}_{i,j \in N}\) is realized then

\[
u_i(\theta_i, p^i, x^*_1, \ldots, x^*_i, \ldots, x^*_n) \geq \nu_i(\theta_i, \tilde{p}^i, x^*_1, \ldots, x^*_i, \ldots x^*_n)
\]

Proof: It is straightforward to check that

\[
u_i(\theta_i, \tilde{p}^i, x^*_1, \ldots, x^*_i, \ldots x^*_n) = \nu_i(\theta_i, p^i, x^*_1, \ldots, x^*_i, \ldots, x^*_n) + c|\{z_1, z_2, \ldots, z_s\}|
\]

A network is SPNE if in the affiliation stage no link proposal is unreciprocated, and in the behavior adoption stage players choose according to Proposition 5. 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 address some criteria of equilibrium selection in networks.

### 5.4.4 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 which networks are pairwise stable, so that once an action profile is chosen, players do not have incentives to increase or decrease their degree. A network is pairwise stable then 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. Because pairwise stability only takes into account link selection, we fix the set of action profiles, \(x\), to formally define the concept for our model, but as it will be shown in Proposition 6, we will adapt the characterization, including the condition that no actor changes her behavior.

**Definition 4** Pairwise stability: Let \(x\) be an action profile. A network \(\{g\}\) generated by \(p\) is pairwise stable if:

---

For different theoretical characterizations of pairwise stability see also Jackson and Watts (2001) and Calvó-Armengol et al. (2009)
1. Suppose $p^i_j = p^j_i = 1$. Consider the network $\tilde{g}$ generated by $\tilde{\mathbf{p}}$ that coincides with $g$ except $\tilde{p}^i_j = \tilde{p}^j_i = 0$, ($i$ and $j$ are not connected) then
   
   $u_i(\theta_i, \mathbf{p}^i, x) \geq u_i(\theta_i, \tilde{\mathbf{p}}^i, x)$ and $u_j(\theta_j, \mathbf{p}^j, x) \geq u_j(\theta_j, \tilde{\mathbf{p}}^j, x)$

2. Suppose $p^i_j = p^j_i = 0$. Consider the network $\tilde{g}$ generated by $\tilde{\mathbf{p}}$ that coincides with $g$ except $\tilde{p}^i_j = \tilde{p}^j_i = 1$, ($i$ and $j$ are connected) then
   
   (a) if $u_i(\theta_i, \mathbf{p}^i, x) \geq u_i(\theta_i, \tilde{\mathbf{p}}^i, x)$ then $u_j(\theta_j, \mathbf{p}^j, x) < u_j(\theta_j, \tilde{\mathbf{p}}^j, x^*)$ or
   
   (b) if $u_j(\theta_j, \mathbf{p}^j, x) \geq u_j(\theta_j, \tilde{\mathbf{p}}^j, x)$ then $u_i(\theta_i, \mathbf{p}^i, x) < u_i(\theta_i, \tilde{\mathbf{p}}^i, x)$

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. Therefore, we select from the set of action profiles those leading to a Nash equilibrium, in the second stage, robust to unilateral and bilateral link changes.

**Definition 5** **Pairwise stable networks:** The pair $(\{g\}, x^*)$ is a pairwise stable Nash equilibrium if

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 proposition characterizes the set of pairs $(\{g\}, x^*)$ which are pairwise stable Nash equilibria in the game $\Gamma$.

**Proposition 6** **Pairwise stable Nash equilibria:** The pair $(\{g\}, x^*)$ is a pairwise stable Nash equilibrium of the game $\Gamma$ if it satisfies the following conditions:

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

(ii) No player wants to change her behavior: Let be $\theta_i = 1$, and $\chi_i$ the number of neighbors of player $i$ playing action 1 in the network $g$. Then,

   (a) if $x^*_i = 1$ then $\chi_i \geq \tau(k_i + 1)$ and $\chi_i - 1 \geq \tau(k_i - 1)$
   
   (b) if $x^*_i = 0$ then $\chi_i + 1 < \tau(k_i + 1)$ and $\chi_i < \tau(k_i - 1)$

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 \(\Gamma\):

\[
\begin{align*}
  u_i(\theta_i, \{\tilde{g}\}, x_1^*, \ldots, x_i^*, \ldots, x_n^*) &> u_i(\theta_i, \{g\}, x_1^*, \ldots, x_i^*, \ldots, x_n^*), \\
u_j(\theta_j, \{\tilde{g}\}, x_1^*, \ldots, x_j^*, \ldots, x_n^*) &> u_j(\theta_j, \{g\}, x_1^*, \ldots, x_j^*, \ldots, x_n^*)
\end{align*}
\]

where it is straightforward to check that

\[
u_i(\theta_i, \{\tilde{g}\}, x_1, \ldots, x_n) = \alpha(k_i + 1) - ck_i > \alpha k_i - c(k_i - 1) = u_i(\theta_i, \{g\}, x_1, \ldots, x_n)
\]
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 \(\Gamma\):

\[
u_i(\theta_i, \{\tilde{g}\}, x_1^*, \ldots, x_i^*, \ldots, x_n^*) > u_i(\theta_i, \{g\}, x_1^*, \ldots, x_i^*, \ldots, x_n^*)
\]

It is straightforward to check that

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

Finally, the proposition 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 \(\chi_i\) and the
threshold functions from Hernández et al. (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 form a link between them but their adopted behavior is uncoordinated, 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, any link to a neighbor who is behaving differently is eliminated. The intuition behind Proposition 6 points to a single argument: for each action profile $x^*$ there is only one network configuration that is a pairwise stable Nash equilibrium.

For the remaining part of this section we introduce a second concept frequently used in network modeling to differentiate outcomes: efficiency. Clearly, a common and natural way to measure efficiency is through the usual notion of Pareto efficiency, where a network is Pareto efficient if no other network leads to better payoffs for all individuals of the society. However, as illustrated in Chapter 4 through the 2-person game, when there are conflicting preferences (i.e. players with different identities interacting in the network) it is not possible to Pareto rank all the equilibria.

For this reason, we use a notion denoted as strong efficiency (Jackson and Wolinsky, 1996) that focuses on the total productivity of a network which, for our case of study, depends on the preferences actors have over the available choices. Strong efficiency allows us to analyze how the players’ incentives, given their individual preferences, align with social efficiency. That is, when the private incentives of individuals to connect with one another lead to network structures that maximize some appropriate measure of social efficiency (Jackson, 2005). In this way we are able to assess whether heterogeneity in preferences between the players, which implies a conflict on what choice they would rather coordinate on, necessarily leads to social exclusion between identities or not.

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

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

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 6 Strong Efficiency:** A pair $(\{g\}, x)$ is strongly efficient in the game $\Gamma$ if $v(\{g\}, x) = \arg\max_{\{g\}, x} v(\{g\}, x)$.

In Proposition 6 we derive from the concept of pairwise stability that only two kind of network configurations can be a pairwise stable Nash equilibrium: (1) a completely

---

5Our measure of strong efficiency focuses on total productivity in the network and does not allow for transfers between the players.
connected structure if the action profile is \textit{specialized}, and (2) a network with two isolated and completely intra-connected components if the action profile is \textit{hybrid}, where each component is specialized in a different action.

Unlike pairwise stability, for which the inclusion of identities is absent, efficiency depends on the \textit{a priori} distribution of identities. We will refer to the distribution of identities as the indicator for the \textit{level of conflict in preferences} in the game, denoted by $\Pi$. This will be particularly useful in our experimental study, presented in the next section. We assume there is a proportion of $\pi_{\theta_i}$ players with identity $\theta_i$, where $\pi_0 + \pi_1 = 1$. 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 $\Pi \in (0, 1)$. The more homogeneous a population is, the lower the level of conflict. Thus, if $\pi_1 = 0$ or $\pi_1 = 1$, then $\Pi = 0$. The more heterogeneous the population is the higher the level of conflict. This means that if $\pi_1 = \pi_0$, then $\Pi = 1$. See Figure 5.1 for an illustration.

![Figure 5.1: The horizontal axis represents the share of players with identity 1 ($\pi_1$) in the population. The vertical axis represents the level of conflict ($\Pi$).](image)

Based on this consideration of conflicting preferences, we want to know what the conditions are, in terms of $\Pi$, for a specialized or a hybrid equilibrium to be strongly efficient. Proposition 7 presents this arguments.

\textbf{Proposition 7 Strongly efficient network:} The strongly efficient configuration of the game $\Gamma$ is the complete network specialized in the prescribed behavior for the majority, for any level of conflict. So that:

(i) if $\Pi = 0$, such that $\pi_1 = 1(0), x^*_i = 1(0)$ and $k_i = (n - 1) \forall i \in N$

(ii) if $0 < \Pi < 1$, such that $\pi_1 > \pi_0 > 0$, $x^*_i = 1$ and $k_i = (n - 1) \forall i \in N$

(iii) if $\Pi = 1$, such that $\pi_1 = \pi_0$, either $x^*_i = 1$ or $x^*_i = 0$, and $k_i = (n - 1) \forall i \in N$

\textbf{Proof:} From Proposition 6 we know that in a pairwise stable configuration every player is connected to all other players choosing the same action as her. This result naturally extends to the efficient network. Thus, 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 $\Pi = 0$ 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 Proposition 6 that such networks are pairwise stable, so that \( k_i = n - 1 \) for all players in the \( F_S \), and \( k_i = n\pi\theta_i \) for players in the \( S_H \).

Consider a distribution of identities such that there are \( n\pi_1 \) players with identity 1 and \( n\pi_0 = n(1 - \pi_1) \) players with identity 0. The aggregate payoffs of the \( F_S \) network are given by:

\[
v(F_S) = \sum_{i=1}^{n\pi_1} \alpha n - c(n-1) + \sum_{i=1}^{n\pi_0} \beta n - c(n-1)
= n[\pi_1(\alpha n - c(n-1)) + (1 - \pi_1)(\beta n - c(n-1))]
\] (5.6)

The aggregate payoffs of the \( S_H \) network are given by:

\[
v(S_H) = \sum_{i=1}^{\pi_1n} \alpha(n - \pi_1(n - 1)) + \sum_{i=1}^{\pi_0n} \alpha\pi_0n - c(\pi_0(n - 1))
= n[\pi_1(\alpha\pi_1n - c(\pi_1(n - 1))) + (1 - \pi_1)(\alpha(n - \pi_1n) - c(n - \pi_1(n - 1)))]
\] (5.7)

where it is straightforward to check that

\[
v(F_S) > v(S_H) \text{ for } \pi_1 \geq \frac{1}{2}
\] (5.8)

The third point is easy to prove under the conditions exposed so far, because if \( \Pi = 1 \) then \( \pi_1 = \frac{1}{2} \), and Equation 5.8 states that under that level of \( \pi_1 \) 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 \( \pi_0 = \pi_1 = \frac{1}{2} \).

The intuition of this Proposition is that when social influence is exerted by a majority, specialization maximizes the total productivity of a network. Specialization leads to this outcome even if the share of each social category in the population is the same. Particularly, when this is the case, socially there is no difference between which behavior players specialize in. 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
5.4 Equilibrium

\[ \pi_{\theta_i} > \frac{(\beta-c)}{(\alpha-c)} > \frac{1}{2} \frac{(\alpha-\beta)}{(\alpha-c)}. \]

Clearly, depending on the values of \( \alpha, \beta \) and \( c \) the strongly efficient configuration may not be payoff dominant for all players, especially for those belonging to the minority. For our experimental game, we particularly focus on a payoff scheme in which the strongly efficient network (i.e. all subjects connect and choose the action preferred by the majority) dominates in payoffs the network where subjects segregate given their individual preferences. With this payoff scheme we are able to focus on the particular problem of how the level of conflict in the network influences the interplay between individual preferences and the benefit of coordinating with those around us.

To illustrate better our payoff scheme, which will be described in more detail in the following section, let’s go 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 vice versa. Furthermore, with the payoff scheme we have chosen, 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.\(^6\)

In conclusion, our identity-based model extends the theoretical work in Chapter 4, particularly addressing the way networks of relationships can emerge from endogenous choices of the players involved. We model situations in which actors can choose what behavior to adopt and also they can decide with whom they want to interact. Thus, in our model actors play a coordination game with multiple partners and can decide who those partners are. We have characterized equilibrium outcomes and integrated notions that help portray different outcomes in networks. The first notion is that of pairwise stable equilibria (PNE). If a network outcome is PNE, no player has incentives to change either her connections (i.e. neither create a relationship with an unconnected player, nor sever a relationship with an existing neighbor) or her behavior. These are situations where players are better off coordinating with all their neighbors on 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.\(^7\)

---

\(^6\)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: \( \alpha n \pi_{\theta_i} - c(\pi_{\theta_i}, n-1) \geq \beta n - c(n-1) \).

\(^7\)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.
The second notion is that of strong efficiency, through which the total productivity of the network (i.e. utilitarian welfare) is measured. The model also points out that depending on the distribution of identities (i.e. the level of conflict in the population) some equilibria dominate others in payoffs. The efficient equilibrium is a network in which all players are connected in one same component and they all choose the behavior prescribed for the majority. Thus, the share of the population that a given identity occupies can determine if players 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 used to test how exclusion of potential partners and inequalities in the benefits acquired by the actors depend on the level of heterogeneity in the network, that is, the level of conflict in preferences.

5.5 The experiment

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.

5.5.1 Experimental game

There are 15 subjects per group in a one-shot interaction. All subjects at the beginning of the interaction are informed about a symbol they are assigned to, either a square or a circle. The two symbols represent the artificially generated social categories to which subjects can belong to. Each participant knows her own and the others’ social category.

The experimental game has two-stages. In the first stage, subjects simultaneously decided to whom they wanted to propose a link (see Figure 5.2). Subjects were 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 (e.g. number 12 always belonged to the social category square). The cost of proposing a link is 2 points 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; the resulting social network (see Figure 5.3). Then, they had to choose one of two options: *up* or *down*. If a subject with identity *square* chooses option *up* she earns 6 points every time she coordinates with a neighbor on that choice. In accordance with the theoretical model we say that this choice represents the action *preferred* by a subject with identity *square*. The same hold symmetrically for subjects with identity *circle* and option *down*. Conversely, when a subject with identity *square* coordinates in option *down* with a neighbor she earns 4 points, for this is the option not preferred by her. The same holds for a subject with identity *circle* and option *up*.

The total number of points earned is calculated with the payoff function in Equation 5.1.\(^8\)

\(^8\)For the chosen payoff scheme we can see that for 14 potential neighbors (there are 15 subjects in each group), the thresholds in Proposition 5 are \(\tau = 6\) and \(\tau = 9\). This means that if a subject can coordinate with 6 others in her preferred choice she has incentives to choose for it. Conversely, if she can coordinate with more than 9 neighbors in her non-preferred option, she has incentives to choose what she dislikes.
expected payoffs in any situation given their assigned identity and the behavior of others. In addition, all subjects received a printed table illustrating the points they can get for any level of connections and any choice in which they coordinated on.

5.5.2 Experimental design and treatments

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, while group composition and the assigned identification number varied. That is, a subject belonged to the same social category for all rounds. The first five were trial rounds.

To evaluate the effect that the level of conflict in preferences has upon outcomes, we used the distribution of identities in the groups as our experimental variable. We implemented three treatments that systematically vary this feature: No conflict (15 majority, 0 minority), Low conflict (12,3) and High conflict (8,7). In all our treatments we kept the social category square as the majority. Our experimental design captures an important mixed-motive social situation from which we derive contrasting hypotheses for the equilibrium selection strategies. On one hand, the individual preference motivation in the identity literature states that if there are artificially induced identities, subjects are more likely to favor their in-group. On the other, the payoff dominant motivation, from the literature on social influence, 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 4a and 5a (4b and 5b) are a result of how identity (social influence) predicts the choices on affiliation and behavior. The hypotheses for the first stage are:

**Hypothesis 4a Identity-dominant affiliation.** The higher the level of conflict the higher the tendency to propose connections only to the in-group.

**Hypothesis 4b Payoff-dominant affiliation.** The level of conflict does not increase the tendency to propose connections only to the in-group, and subjects connect also to out-group

The affiliation hypotheses 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

---

9From our model we know there are two classes of outcomes that are pairwise stable: (1) two complete but separate components, each specialized in a different action, and (2) a complete network specialized in one action. For our mixed motive experimental design, with the payoff schemes chosen, if identity is dominant subjects would segregate between circles and squares, each choosing their favorite option. If so, a subject earns 90 points in No conflict, 72 points if square and 18 points if circle in Low conflict, and 48 points if square and 42 points if circle in High conflict. If payoffs instead of identity are dominant, the optimal choice is to integrate and specialize. The pairwise stable and strong efficient equilibrium for our payoff scheme is the complete network specialized in choosing up: the prescribed behavior for the majority. In this case, for all treatments a square earns 90 points and a circle earns 60 points.
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 4b, and segregation is predicted for treatments with conflicting preferences by Hypothesis 4a.

**Hypothesis 5a** *Identity-dominant behavior*. The higher the level of conflict the more likely subjects will choose the option they prefer given their identity.

**Hypothesis 5b** *Payoff-dominant behavior*. The level of conflict will have no effect and subjects are more likely to all choose the same option regardless of their identity.

The behavior adoption hypotheses state that if identity is more salient than social influence (i.e. payoffs), in treatments with a positive level of conflict, subjects are more likely to choose the behavior each prefers. Otherwise, subjects all specialize in the same behavior. For instance, if the equilibrium chosen is the strongly efficient then 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 for equilibrium selection. Notice, nevertheless, that as argued by Camerer (2003), it is unlikely that equilibrium is reached instantaneously in one-shot games. A more useful perspective is 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. The following are the hypotheses on equilibrium derived from our game theoretic model:

**Hypothesis 6** *Reciprocity*. The higher the number of one-shot interactions subjects are part of, the smaller the difference between links proposed and links formed.

This prediction is derived for the affiliation stage of our network game from the backward induction process; subgame perfect Nash equilibria. Finally, the hypothesis on connectivity is derived from our modeling of equilibrium selection and the pairwise stability:

**Hypothesis 7** *Connectivity*. The higher the number of one-shot interactions subjects are part of, the more likely subjects choosing the same action will be neighbors.

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. Regardless, these hypotheses predict that networks will
tend to be more densely connected along time, leading towards the pairwise stable predicted configurations.

5.5.3 Experimental procedures, data and methods

The experiment was conducted in the Laboratory of Experimental Economics (LINEEX) at the University of Valencia in November 2012. Subjects interacted for 25 rounds through computer terminals and the experiment was programmed using z-Tree (Fischbacher, 2007). Upon arrival subjects were randomly seated in the laboratory. At the beginning of the experiment instructions were read out loud to all subjects to guarantee that they all received the same information (see Appendix C). 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.

Subjects were recruited through online recruitment systems in the campus of social sciences of the University of Valencia (Spain). In total 120 subjects participated in three sessions, each lasting between 90 and 120 minutes, one for each treatment (No, Low and High Conflict). There were 30, 45 and 45 participants in each session, respectively, 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 used to test the hypotheses presented above, and 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, the number of proposals she sent to the in-group was divided by her total number of proposals sent. In-group favoritism ranges between 0 and 1, where 1 (0) means that all proposals were sent to the in-group (out-group). A value of 0.5 denoted equal preferences for sending proposals to both the in-group and the out-group. This measure is only used for treatments High and Low because in No Conflict there is no out-group.

**Reciprocation.** A subject’s number of reciprocated link proposals divided by a subject’s total number of proposals. Reciprocation had a maximum of 1 (0) when all proposals were reciprocated (rejected).

**Connectivity.** 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. A value of 1 expressed maximum connectivity. 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.\(^\text{10}\)

\(^{10}\)The *connectivity* measure plays an equivalent role to pairwise stability for the results of our experiment. There is virtually no variance in the choices subjects made, as illustrated in Table 5.1: 99.3% chose their favorite action and 99.4% connected only with in-group.
Analytical strategy. The data structure at hand did not permit standard ordinary least square regression modeling, for it is based 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. Note also that interdependencies within experimental sessions are not part of the analyses.

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. Three separate models were run for in-group favoritism, reciprocation and connectivity.

5.6 Results

The data show that nearly all choices corresponded with the subjects’ preference. We observed that in 99.3 percent of the cases 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 5.1 presents an overview of the proposals sent and reciprocated (i.e. the realized connections) for the different experiment treatments and groups. In-group favoritism was virtually identical in the conflict treatments. Stronger in-group favoritism related to increased reciprocation (Pearson’s correlation coefficient: $r = .694, p < .001$), which in turn was associated with greater connectivity ($r = .687, p < .001$). We below describe how these choices confirm the identity-dominant hypotheses.

Hypothesis 4a expected that higher level of conflict would lead to greater favoritism for in-group proposals. The alternative Hypothesis 4b stated that no such effect would occur. Table 5.2 presents the results from the mixed-effects regression models. The constant of 0.84 indicates that in-group favoritism was generally high: putting aside all other variables (treatments, group membership and development over periods), it could be predicted that subjects send proposals to members from their own group in 84 percent of the cases. However, there is no significant difference between treatments for in-group favoritism. For both Low Conflict and High Conflict, subjects’ proposals to their in-group compared to out-group is virtually the same. These results were not considered in the proposed hypotheses, and both H4a and H4b are rejected, suggesting there is no effect of conflict in the choices of affiliation. As long as conflict exists, subjects proposed mainly to in-
### Experimental condition

<table>
<thead>
<tr>
<th>Group</th>
<th>Experimental condition</th>
<th>No Conflict</th>
<th>Low Conflict</th>
<th>High Conflict</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>Majority</td>
<td>Proposals (^a)</td>
<td>12.87</td>
<td>2.47</td>
<td>9.82</td>
</tr>
<tr>
<td>Minority</td>
<td></td>
<td>n/a</td>
<td>n/a</td>
<td>1.96</td>
</tr>
<tr>
<td>Majority</td>
<td>to in-group</td>
<td>11.95</td>
<td>2.90</td>
<td>8.79</td>
</tr>
<tr>
<td>Minority</td>
<td></td>
<td>n/a</td>
<td>n/a</td>
<td>1.92</td>
</tr>
<tr>
<td>Majority</td>
<td>Connections (^a)</td>
<td>11.95</td>
<td>2.90</td>
<td>8.79</td>
</tr>
<tr>
<td>Minority</td>
<td></td>
<td>n/a</td>
<td>n/a</td>
<td>1.92</td>
</tr>
<tr>
<td>Majority</td>
<td>to in-group</td>
<td>11.95</td>
<td>2.90</td>
<td>8.79</td>
</tr>
<tr>
<td>Minority</td>
<td></td>
<td>n/a</td>
<td>n/a</td>
<td>1.92</td>
</tr>
<tr>
<td>Majority</td>
<td>to out-group</td>
<td>1.96</td>
<td>0.24</td>
<td>0.28</td>
</tr>
<tr>
<td>Minority</td>
<td></td>
<td>5.97</td>
<td>0.17</td>
<td>0.28</td>
</tr>
<tr>
<td>Majority</td>
<td>In-group favoritism (^b)</td>
<td>1.00</td>
<td>0.00</td>
<td>0.98</td>
</tr>
<tr>
<td>Minority</td>
<td></td>
<td>n/a</td>
<td>n/a</td>
<td>0.83</td>
</tr>
<tr>
<td>Majority</td>
<td>Reciprocation (^b)</td>
<td>0.92</td>
<td>0.12</td>
<td>0.87</td>
</tr>
<tr>
<td>Minority</td>
<td></td>
<td>n/a</td>
<td>n/a</td>
<td>0.84</td>
</tr>
<tr>
<td>Majority</td>
<td>Connectivity (^b)</td>
<td>0.85</td>
<td>0.21</td>
<td>0.80</td>
</tr>
<tr>
<td>Minority</td>
<td></td>
<td>n/a</td>
<td>n/a</td>
<td>0.96</td>
</tr>
</tbody>
</table>

\(^a\) Means and standard deviations for proposals and connections represent absolute numbers.

\(^b\) Means and standard deviations for in-group favoritism, reciprocation and connectivity represent relative shares (percentages).

**Table 5.1:** Subjects’ proposals and connections within and between groups (across all periods).
Hypotheses on behavior adoption stated that if subjects were more influenced by their identities the higher the level of conflict Hypothesis 5a the more likely they were to behave as prescribed for their social category. Alternatively Hypothesis 5b 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 treatments. Thus, the evidence suggests that regardless of the level of conflict, subjects’ behavior is influenced by identities above social pressure. Result 1 summarizes these findings:

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

Hypothesis 6 expected learning and thus increasing 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 with every additional interaction. By pursuing the prescribed behavior for their social category, subjects segregate, and the conflicting aspect of the interaction is put aside.

The two components in the network appear as if they were two isolated populations. Once subjects end up in a network connected to others who share their same identity,
social influence takes a relevant role again by means of reciprocation, and subjects start aiming to connect with all those around them. Thus, 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 7 predicted that subjects will coordinate more along time so that the links proposed are formed, and that they 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 connectivity increased with every additional interaction. It was reasonable to assume that the learning curve for reciprocation and connectivity 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 connectivity did not increase significantly anymore in later periods, namely after period 10, suggesting a curvilinear learning effect. These findings are illustrated in Figure 5.4 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 treatments, further supporting our assumptions. Besides differences between treatments 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 connectivity than those in the minority group. This effect occurred net of the different treatments.

The difference between majority and minority group persisted throughout the entire experiment, but differences became smaller toward high numbers of one-shot interactions. We interpret this to be mainly due to the learning effect in the majority group. While on average subjects in the minority reached maximum connectivity of 1 in period 4, subjects in the majority reached their maximum of 0.95 only in period 20.

**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.*

### 5.7 Discussion

In this article we have argued that the interplay between individual preferences and social influence can decisively affect outcomes of how people relate to each other and what choices
Figure 5.4: Predictive margins by period.
they make in terms of their behavior.

Our model indicated that the choice an actor makes about what behavior to adopt depends 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 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 is exerted.

Given the multiplicity of equilibrium outcomes we make use of different equilibrium selection criteria. Pairwise stable configurations are 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 everyone is behaving in the same way, or a network separated into two completely intra-connected components, where actors in each component behave the same but not between components.

Finally, we ranked the social efficiency of the resulting networks. This means that for both the majority and the minority, the payoff maximizing strategy is to connect together, as long as all players choose the same. Moreover, the network where all actors are connected and their behavior is the same gives the highest total productivity when the behavior chosen is that of the majority. We particularly choose a payoff scheme for which the strong efficient network dominates in payoffs the pairwise stable configuration formed by two separate components. By doing so, we have been able to focus on the interplay between individual differences and the benefits of choosing alike with others, in settings of conflicting preferences.

To test our theory we designed an experimental study in which we varied the composition of the population for three treatments: 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 conflicting preferences about what behavior to adopt, individual identities are more salient than social influence. Therefore, networks segregate into two components, each formed by subjects with the same identity and all choose the behavior they prefer given their identity. This first result reinforces the categorization argument of identity theory showing how identities can be so strong that are used to 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 reduces the total productivity of society in an important way.

Our second empirical result states that once segregation emerges, so that identities are not in conflict anymore, social influence becomes more salient so that actors aim to connect completely within their component. The conflict in preferences makes the payoff domi-
nant structure unreachable but once in the segregated configuration, it leads to the payoff dominant structure for the component. This points to the tension between stability and efficiency that has been so relevant and pervasive in network studies (Jackson and Wolinsky, 1996), but introduces the effect of identities in it, showing 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, the failing of coordination was not present in the No Conflict treatment. When all subjects were an absolute majority and group size was the largest, they did not show the same limitations in maximizing the complementarities of their social connections. Thus, suggesting that it is the presence of an out-group minority and not the size of the in-group what promotes the difficulties to connect.

This result complements the existing work on in-group bias in identity theory, when identification is experimentally induced (i.e. minimal group paradigm). In-group bias is significantly observed, even for cases where there is a majority and a minority (Leonardelli and Brewer, 2001). Our results complement these findings 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.

Given the close relation between our study and Bojanowski and Buskens (2011), here we discuss in more detail how the results of our experimental study complement and extend theirs. Our findings suggest that groups will segregate in the emerging network when actors can choose with whom to form costly relationships. This is not always the case in Bojanowski and Buskens (2011), while our experimental subjects segregate in every opportunity they have to interact. This difference might be due to particular aspects of our payoff function and the linear cost assumption in our model compared to the different variations in Bojanowski and Buskens (2011) denoted as complementarity and substitutability. In particular, while subjects in our model only gain from a relationship if they coordinate in the behavior with their neighbor, in Bojanowski and Buskens (2011) actors gain benefit, in each relationship, by choosing their preferred (native) behavior regardless of coordination. This can also explain why our experiment shows that segregation between groups is linked with groups choosing the behavior associated with their type.

In both works we find that when networks segregate, they tend to be densely connected. However, a main important difference between the studies is that we observe that for experimental subjects their individual types are more salient than payoffs. Therefore, irrespective of the relative group size (i.e. majority vs. minority) segregation occurs in our experimental treatments with conflict. While Bojanowski and Buskens (2011) find that the smaller the minority relative to the majority, the more likely integration will occur. Our experimental findings suggest that for our model this is caused by the low
reciprocity that subjects in the majority give to link proposals for the minority in the first rounds, specially in the High Conflict treatment.

Some limitations of our work warrant further discussion. Compared to other works on identities assuming that individuals can choose their individual identity and not only their behavior (Akerlof and Kranton, 2000), we model social categories as fixed. Our 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.

Session effects, where observations across subjects in a session might exhibit more correlation than observations across subjects in different sessions, are a common problem in laboratory experiments. A higher number of sessions would have been desirable to be able to assess the potential bias that session effects may have caused in our study. However, data collection was costly, also because in our study each group involved a particular high number of subjects (15 subjects), so that a study design with much more groups (and thus sessions) was not feasible. While we cannot rule out session-based correlations across subjects completely, we are confident that bias was limited, as selection into the different groups was completely at random and subjects had no personal contact at any time point during the experiment. Future research may incorporate higher number of sessions and thereby control for these context effects.

An important line that can further extend our work could look at the inclusion of communication between subjects before they make their choices. Pre-play communication has been shown to facilitate coordination and equilibrium play (Farrell and Rabin, 1996). Communication could work in two different directions and thus its study has the chance to enhance our understanding of the interplay between individual preferences and social influence. On the one hand, it could serve as a device that helps players signal their intention to maximize payoffs by integrating with their out-group. One the other, communication could reinforce in-group bias and segregation, by facilitating higher levels of connectivity with between players sharing common identities.