Journal of Dynamics and Games
Volume **3**, Number **2**, April **2016**

doi:10.3934/jdg.2016010

pp. **191–216**

CONFLICT AND SEGREGATION IN NETWORKS: AN EXPERIMENT ON THE INTERPLAY BETWEEN INDIVIDUAL PREFERENCES AND SOCIAL INFLUENCE

Lea Ellwardt

Institute of Sociology and Social Psychology
University of Cologne, Cologne, Germany
Albertus Magnus Platz
50923 Cologne, Germany

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

ERI-CES, Departamento de Análisis Económico, Universitat de Valencia
Avenida Los Naranjos s/n, Campus de Tarongers
Valencia, 46022, Spain

(Communicated by Filomena García)

Abstract. 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 where not all have a preference for 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 have the same preferences. This leads to two undesirable outcomes: network segregation and social inefficiency. The same force that helps people individually, hurts society.

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 behaviour and whether to establish relationships with others or not. For instance, when choosing our friends [32] or neighbours [36] individual preferences are a strong determinant of how we make such decisions. But also, the social influence peers exercise on human behaviour is enormous [23], affecting whether people act in alignment with those they relate to or not [33]. Examples of social influence range from which products we buy [17], to whether we engage in criminal activities or not [6], to our participation in collective action [20]. Our aim in this paper 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.

2010 *Mathematics Subject Classification.* Primary: 91A90; Secondary: 91A40.

Key words and phrases. Heterogeneity, social networks, formation, equilibrium selection.

We thank the Laboratory for Research in Experimental Economics of the University of Valencia (LINEEX) for their help running the experiment. We also thank Arno Riedl, Tom Snijders and Vincent Buskens for valuable comments in previous versions of this manuscript.

* Corresponding author: Penélope Hernández.

One of the most prominent theoretical tools to study the effect individual preferences have on the way people behave is *identity theory* [38, 1]. From the perspective of identity theory a person's sense of self, her identity, is composed of three elements. First, *categorisation*, putting ourselves and others into social categories (e.g. being an Orthodox christian, a female, a policeman). Second, *identification*, the process we use to associate ourselves with certain groups. The group 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 particular behaviours prescribed for them. We refer to this prescribed behaviour 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 living up to the norms set by their social categories are unhappy, so they tend to change their decisions to meet their standards [4].

On the other hand, *strategic interaction in networks* is a leading research program studying how we make our choices influenced by our social relationships. 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 [43]. These interactions are known as coordination games with strategic complementarities, where a person's incentives to adopt a given behaviour increase as more of those around her make the same choice. The underlying mechanism from social influence is that people perceive coordinating with the behaviour of others as beneficial for them. As a result, people are more likely to adopt a given behaviour depending on how others behave, even if such a behaviour is not the one prescribed for their identity [21].

The existing research on these two lines of work has illustrated ways in which identities *or* social influence affect our relationships and our behaviour. 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 analyse a formal model where actors choose with whom to interact and which behaviour to adopt, and experimentally tests the model by varying the way identities and social influence interact. 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 behaviour is made once the structure of relationships has been formed. There is one behaviour prescribed to each social category, so that the preference of an individual is to adopt the behaviour that corresponds 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, in order to elucidate 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. Actors are better off aligning their behaviour to that of those around them. However, their identities introduce a conflict about which behaviour each prefers to adopt. Thus, we model the interplay between identities and social influence in a context of *conflicting preferences*. We design 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 by everyone. We vary the relative size of the social categories, which changes the intensity of the conflict in preferences between subjects.

The remainder of this paper is assembled as follows: In section 2 we describe our theoretical framework. The game theoretic model is presented in Section 3. Section 4 analyses the network structures that emerge from the interactions of actors belonging to different social categories. In section 5 we describe the experimental study. Section 6 presents the main results of our experiment. We conclude with a discussion of the implications and limitations of the study in Section 7.

2. **Theoretical framework: Identities and social influence.** Our theoretical framework builds on two lines of work examining how relationships and behaviour emerge from actors' individual preferences and the influence from those around them: *identity theory* (originating from Psychology but recently increasingly adopted in Economics, see [4]) and *strategic interaction in networks* (from Economics). Our study integrates both lines of research for interactions with *strategic complementarities*.

Research on the theory of identities was initiated in Psychology [38, 40], mainly focusing on the effects that the social context has on group processes and inter-group 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/discriminate against each other. A consistent finding in identity theory is that people favour their in-group relative to out-groups (i.e. in-group bias), because people desire a positive and secure self-concept, so they think of theirs as good groups.

The argument of in-group bias has been widely supported by experimental research on identities [8]. 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 (e.g. preferences 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 [9, 34], 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 behaviour of others. Participants in these experiments could not benefit or lose in any way from their point allocation strategy, and in some experiments points did not even carry any value at all [39]. 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 [1]. The application of their model has been found useful to explain gender discrimination [1], education [2], and contract theory [3]. Experimental research has also included identity as part of the analysis, addressing the limitation in the psychological approach by taking into account monetary stakes [7, 18, 12, 13]. Particularly, [13] 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 modelling of identities in economics provides insight into broad patterns of social behaviour, 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 processes, and have proven to be necessary for understanding the way others influence our behaviour.

Research on networks introduced the strategic behaviour of people into the analysis of social influence by modelling interactions as games (for surveys of the literature see [19, 41, 24]). Network games model the way individuals behave as a function of the actions of their neighbours. In settings where individuals are better off the more of their neighbours behave as they do but where there are at least two possible behaviours, influence is captured by thresholds functions [20, 17]. For instance, when a person is deciding whether to acquire a specific technology or not, if more than a given number of her neighbours (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 in the same vein includes [14], [28], [44], [33], and [30]. A persistent finding in the theoretical modelling 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 behaviour 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 modelling social relationships as endogenous decisions actors make [27]. 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 with this aim include [27], [5], [26] and [35]. 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 [26]. 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 aspect 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 [21]. In their model the authors address the effect of heterogeneity on 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 to whom they relate. 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 [23] and show that is it more likely to have friends of the same race [31] or gender [42]. By modelling both stages we can study how the level of conflicting preferences influences people's behaviour, given the interplay between individual preferences and social influence.

3. **The model.**

In this section we present our model of network interactions where players have identities, each identity is associated with a behaviour that gives it higher payoffs than the other, and the identities and behaviour 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 behaviour 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 $\mathbf{p_i}$ in $\{0, 1\}^N$, where $p_{ij} = 1$ means that player i proposes a link to j, and $p_{ij} = 0$ that no proposal is made. We assume $p_{ii} = 0$. Only if $p_{ij} = p_{ji} = 1$, 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. 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 neighbours she has, $k_i(g) = \{j : g_{ij} = 1\} \ \forall \ j \neq i$, with cardinality k_i.

In the second stage of the game: *behaviour 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 neighbours 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 behavioural prescription expressed through the linear payoff function, u_i, that strategically depends on the choices made by connected players (we denote $x_{k_i}(g)$ as the vector of actions taken by i's neighbours), their identities and proposed links in the first stage, as follows:

$$u_i(\theta_i, \mathbf{p}, x_i, x_{k_i}(g)) = \lambda_{x_i}^{\theta_i} \left(\sum_{j=1}^{n} p_{ij} p_{ji} (1 - (x_i - x_j)^2) \right) - c \sum_{j=1}^{n} p_{ij}, \tag{1}$$

where $I_{\{x_j = x_i\}}$ is the indicator function of those neighbours choosing the same action as player i. The parameter λ is defined by $\lambda_{x_i}^{\theta_i} = \alpha$ when a player chooses the action prescribed for her identity, and $\lambda_{x_i}^{\theta_i} = \beta$ otherwise. The cost of proposing a link is $c > 0$, and the relation between the parameters in the model is $0 < c < \beta < \alpha$. Note that the cost of proposing a link, c, is paid independently of whether a connection is formed or not. We assume c to be lower than β. Otherwise, the only outcome is the empty network because the benefit of coordinating one's behaviour to that of a neighbour would not be enough to cover the cost of affiliation.

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 $\mathbf{p}$. 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:

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

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 neighbours do not change her payoff.

4. **Equilibrium characterisation.**

In this section we provide the Nash equilibrium characterisation for our network game, NE(Γ). To do this we follow [21], who model network games in fixed networks. We extend their analysis with the characterisation of the subgame perfect Nash equilibria of the two-stage network game.[1]

4.1. **Network categorisation.**

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 (*specialised*) 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 specialised, S_S, where all players coordinate on the same action, which is their preferred choice; (ii) frustrated specialised, 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 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.

4.2. **Nash equilibrium.**

Once the network is realised, the results in [21] for fixed networks are applicable to our case. There are two threshold functions when players have conflicting preferences. A function $\underline{\tau}(k_i)$ that represents the minimum number of i's neighbours choosing the action she likes, for her to choose her favourite action as a best response, and the threshold $\overline{\tau}(k_i)$ representing the maximum number of

---

[1] Notice that along the analysis we can assume without loss of generality a normalisation 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 realised, 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: $[u_i(1, \mathbf{p_i}, 1, x_{N_i}(g)) - c\mathbf{p_i}] - [u_i(1, \mathbf{p_i}, 0, x_{N_i}(g)) - c\mathbf{p_i}] = u_i(1, \mathbf{p_i}, 1, x_{N_i}(g)) - u_i(1, \mathbf{p_i}, 0, x_{N_i}(g))$. Therefore, this cost is cancelled on both sides of the computation.

[2] We denote action profiles as satisfactory or frustrated following the arguments in [1]. When a player adopts the behaviour prescribed for her identity, this reinforces who she is. However, anyone who chooses the non-prescribed behaviour suffers a loss in her identity, entailing a reduction in her utility. That is why $\alpha > \beta$.

neighbours choosing the non-favourite action so that i's best response is still to adopt the behaviour she likes. If one more of her neighbours chooses the non-favourite action, i's best response is to adopt her disliked option. Proposition 1 shows this, where the number of i's neighbours choosing action 1 is χ_i and the number of her neighbours choosing action 0 is $k_i - \chi_i$.[3]

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

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

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

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

The intuition behind Proposition 1 is that a player i wants to coordinate with the highest number of neighbours 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 behaviour ($\overline{\tau}(k_i) > \underline{\tau}(k_i)$), compared to an analysis ignoring identities.

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 characterisation 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 neighbours in $\{g\}$, it could be because she proposed a link to only her k_i neighbours, or because she proposed links to those and even more players; who did not proposed a link back to i.

Let us now define the SPNE in the game Γ. In the first stage of the sequential game Γ, players make connection proposals $\mathbf{p} = (\mathbf{p_1}, \ldots, \mathbf{p_N})$ that result in a network g. Then, in the second stage they play the game Γ_g where players play actions $\{0, 1\}$. Consider the Nash Equilibrium action profile $x^* = (x_1^*, \ldots, x_N^*)$ in the game Γ_g with utility function 1. Following [21], for any the network g, the set of Nash equilibria $x^* = (x_1^*, \ldots, x_N^*)$ may be no unique and vice-versa; fixing a Nash equilibrium $x^* = (x_1^*, \ldots, x_N^*)$ there may exist more than one network g such that $x^* = (x_1^*, \ldots, x_N^*)$ is an equilibrium payoff of Γ_g. Moreover, fixing an action profile x, consider a n-players game with a proposal profile of connections $\{p_{ij}\}_{i,j \in N}$ and Equation 1 as the corresponding utility function. Denote by Γ_x such a game. By

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

backward induction, a Subgame Perfect Nash Equilibrium of the game Γ is defined as a pair $\{p^*, x^*\}$ such that

- x^* is a Nash Equilibrium in Γ_g when p^* generates the network g, and
- p^* is a Nash equilibrium of the game Γ_{x^*}

Next we characterise the best response in the first stage of the game Γ for a player i. Let us fix x^* and $\{\mathbf{p_{-i}}\}$ when there is a positive cost to propose links to other members, even when the link is not realised because it was not reciprocated.

Lemma 4.1. *$p_{ij} = 0$ if and only if either $x_i^* \neq x_j^*$ or $p_{ji} = 0$, and $p_{ij} = 1$ if and only if $x_i^* = x_j^*$ and $p_{ji} = 1$.*

Proof. If $x_i^* \neq x_j^*$ player i would not get any positive payoff form her link with j, so her best response is $p_{ij} = 0$.

If $p_{ji} = 0$, the link between i and j will never exist and therefore there is no possibility for i to get any positive payoff with j, so i's best response is $p_{ij} = 0$.

If $p_{ji} = 1$ and $x_i^* = x_j^*$, player i's best response is $p_{ij} = 1$ if and only if

$$\begin{aligned} u_i(x^*, \mathbf{p_i} : p_{ij} = 1, \mathbf{p_j} : p_{ji} = 1, \mathbf{p_{-\{i,j\}}}; \theta_i) \geq \\ u_i(x^*, \mathbf{p_i} : p_{ij} = 0, \mathbf{p_j} : p_{ji} = 1, \mathbf{p_{-\{i,j\}}}; \theta_i) \end{aligned} \tag{6}$$

Since i and j are coordinating their actions, they get α or β if connected and 0 otherwise. □

The above Lemma states how the set of subgame perfect equilibria of Γ shrinks and actually the equilibrium outcomes are dense networks where nodes are linked if they match their actions as well. Notice that the subgame perfect equilibrium captures the same idea as pairwise stability [27] [4] for the formation game. Therefore, the concept of SPNE embodies the two stability concepts referring to action and formation activities: Nash equilibrium in Γ_g and pairwise stability in Γ_x, respectively.

4.4. **Efficiency.** In this section we study a concept frequently used in network modelling 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 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 [27] 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 analyse 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 maximise some appropriate measure of social efficiency [22] [5].

[4]For different theoretical characterisations of pairwise stability see also [25] and [10].

[5]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:

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

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

We derive from the concept of subgame perfection that only two kind of network configurations can conform a Subgame Perfect Nash Equilibrium: (1) a completely connected structure if the action profile is *specialised*, and (2) a network with two isolated and completely intra-connected components if the action profile is *hybrid*, where each component is specialised in a different action.

Unlike SPNE, for which the inclusion of identities is absent, efficiency depends on the *a priori* distribution of identities. We will refer to the distribution of identities as the indicator for the *level of conflict in preferences* in the game, denoted by Π. This will be particularly useful in our experimental study, presented in the next section. We assume there is a proportion of π_{θ_i} players with identity θ_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 $\Pi = 1$.

Based on this consideration of conflicting preferences, we want to know what are the conditions, in terms of Π, for players to choose a specialised or a hybrid equilibrium. Proposition 2 presents this arguments.

Proposition 2. *The strongly efficient configuration of the game Γ is the complete network specialised in the prescribed behaviour 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 $k_i = (n-1)\ \forall\ i \in N$

Proof. As we know, in a subgame perfect 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 behaviour are connected dominates in payoffs any network with fewer connections. Moreover, such a network will rank the highest if all players are choosing the behaviour they like. For the case of $\Pi = 0$ this is the Satisfactory specialised (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 specialised 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 4.1 that such networks are subgame perfect Nash equilibria, 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:

$$\begin{aligned} v(F_S) &= \sum_{i=1}^{\pi_1 n} \alpha n - c(n-1) + \sum_{i=1}^{\pi_0 n} \beta n - c(n-1) \\ &= n[\pi_1(\alpha n - c(n-1)) + (1-\pi_1)(\beta n - c(n-1))] \end{aligned} \tag{7}$$

The aggregate payoffs of the S_H network are given by:

$$\begin{aligned} v(S_H) = &\sum_{i=1}^{\pi_1 n} \alpha(\pi_1 n) - c(\pi_1(n-1)) + \sum_{i=1}^{\pi_0 n} \alpha\pi_0 n - c(\pi_0(n-1)) \\ &= n[\pi_1(\alpha\pi_1 n - c(\pi_1(n-1))) + (1-\pi_1)(\alpha(n-\pi_1 n) - c(n-\pi_1(n-1)))] \end{aligned} \tag{8}$$

where it is straightforward to check that

$$v(F_S) > v(S_H) \text{ for } \pi_1 \geq \frac{1}{2} \tag{9}$$

The third point is easy to prove under the conditions exposed so far, if $\Pi = 1$ then $\pi_1 = \frac{1}{2}$, and Equation 9 states that under that level of π_1 the aggregate generated profit in a Frustrated Specialised 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, it is socially better for the minority to choose against their identity and increases their benefits from the complementarities of coordinating with more neighbours. Specialisation is socially better 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 on which behaviour players specialise in. Nonetheless, this is the aggregate welfare and for cases with a strict majority it is not always the case that the minority maximises individual payoffs by following this strategy. In fact, a player i from the minority gets higher payoffs in the satisfactory specialised network in which each component is completely connected as long as $\pi_{\theta_i} \leq \frac{(\beta-c)}{(\alpha-c)}$.[6]

In conclusion, our identity-based model states that players are better off coordinating with all their neighbours in the same behaviour, 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* neighbour. The model also points that depending on the distribution of identities (i.e. the level of conflict in the population) some equilibria dominate others. The socially efficient equilibrium is a network in which all players are connected in one same component and they all choose the behaviour 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, sacrificing their identity-based preferences for their social interaction benefit. In the next section we describe the experimental study we used to test our theory.

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

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 behaviour.

5.1. **Experimental game.** There are 15 subjects per group in a one-shot interaction. All subject 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. 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. Then, they had to choose an action *up* or *down*. Up (down) gives 6 points to a subject with identity *square* (*circle*) for every neighbour she coordinates with. Down (up) gives her 4 points. These choices represent the identity-prescribed behaviour 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 assigned identity and the behaviour 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.

5.2. **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. Therefore, for all treatments the socially efficient and individually payoff-dominant outcome is the complete network specialised in choosing *up*, the prescribed behaviour for 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 favour 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 [7]. Thus, hypotheses 1a and 2a (1b and 2b) are a result of how identity (social influence) predicts equilibrium selection. The hypotheses for the first stage are:

Hypothesis 1a. (Identity-dominant affiliation) The higher the 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 maximise 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 behaviour) The higher the level of conflict the more likely subjects will adopt the behaviour they prefer as prescribed by their social category.

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

The behaviour adoption hypotheses state that if identity is more salient than social influence (i.e. payoffs), in treatments with positive levels of conflict, subjects are more likely to choose the behaviour each *prefers*. Otherwise, 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 specialised 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 [11], 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, optimisation, and equilibration (or learning) work. The following are the hypotheses on equilibrium derived from our game theoretic model:

Hypothesis 3. (Reciprocity) 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 connectivity is derived from our modelling of equilibrium:

---

[7] 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 favourite 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 specialise. The SPNE and strong efficient equilibrium for our payoff scheme is the complete network specialised in choosing up; the prescribed behaviour for the majority. In this case, for all treatments a square earns 90 points and a circle earns 60 points.

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

If learning is manifested along the repeated interactions, subjects choosing the same behaviour are more likely to be connected, regardless of whether identities or social influence motivate their behaviour. 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 dense along time, leading towards the SPNE predicted configurations.

5.3. **Experimental procedures, data and methods.** The experiment was conducted in the Laboratory of Experimental Economics (LINEEX) at the University of Valencia. Subjects interacted for 25 rounds through computer terminals and the experiment was programmed using z-Tree [16]. 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 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.

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 their identities are different.

In-group favouritism. To assess a subject's favouritism 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 favouritism 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.

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 realised 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.

Analytical strategy. The data structure at hand did not permit standard ordinary least square regression modelling, 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 behavioural tendencies.

Multilevel regression modelling is a methodology for the analysis of complex data patterns with a focus on nesting [37]. 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 behaviour. Three separate models were run for in-group favouritism, reciprocation and connectivity.

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 behaviour 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 behaviour. Table 1 presents an overview of the proposals sent and reciprocated (i.e. the realised connections) for the different experiment treatments and groups. In-group favouritism was virtually identical in the conflict treatments. Stronger in-group favouritism 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 1a expected that higher level of conflict would lead to greater favouritism for in-group proposals. The alternative Hypothesis 1b stated that no such effect would occur. Table 2 presents the results from the mixed-effects regression models. The constant of 0.84 indicates that in-group favouritism was generally high: putting aside all other variables (treatments, group membership and development over periods), it could be predicted that subjects sent proposals to members from their own group in 84 percent of the cases. However, there is no significant difference between treatments for in-group favouritism. 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 H1a and H1b are rejected, suggesting there is no effect of conflict in the choices of affiliation. As long as conflict exists, subjects proposed mainly to in-group.

Hypotheses on behaviour adoption stated that if subjects were more influenced by their identities the higher the level of conflict Hypothesis 2a the more likely they were to behave as prescribed for their social category. Alternatively Hypothesis 2b expected subjects to be more influenced by their social context choosing the behaviour prescribed for the majority. As mentioned above, 99.3 percent of the choices corresponded to the behaviour prescribed for each subject's individual preference. 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' behaviour is influenced by identities above social pressure. Result 1 summarises these findings:

		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 favouritism [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
Connectivity [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

[a] Absolute numbers.

[b] Relative shares (percentages).

Table 1. Subjects' proposals and connections within and between groups (across all periods).

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

Hypothesis 3 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 behaviour 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, maximising the complementarities of coordinating with their neighbours.

Similarly to the latter hypothesis, Hypothesis 4 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

	Model A In-group favouritism		Model B Reciprocation		Model C Connectivity	
	B	SE	B	SE	B	SE
Low Conf.[a]	n/a	n/a	-0.06^{*}	(0.02)	-0.05	(0.03)
High Conf.	0.03	(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	0.84^{***}	(0.03)	0.79^{***}	(0.02)	0.76^{***}	(0.03)
$N_{observations}$	1,799		2,399		2,399	
$N_{individuals}$	90		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	1,527.7		1,783.91		1,489.50	

[a] For Model A the reference treatment is Low Conflict and for B and C it is No Conflict

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

TABLE 2. Mixed-effects regression models on favouritism, reciprocation and 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 1 summarised in the next result:

Result 2. In the presence of conflicting preferences, when segregation arises between social categories, subjects aim to maximise 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 a maximum of 0.95 only in period 20.

Result 3. In the presence of conflicting preferences, when segregation arises between social categories, being in the minority facilitates coordination and stability.

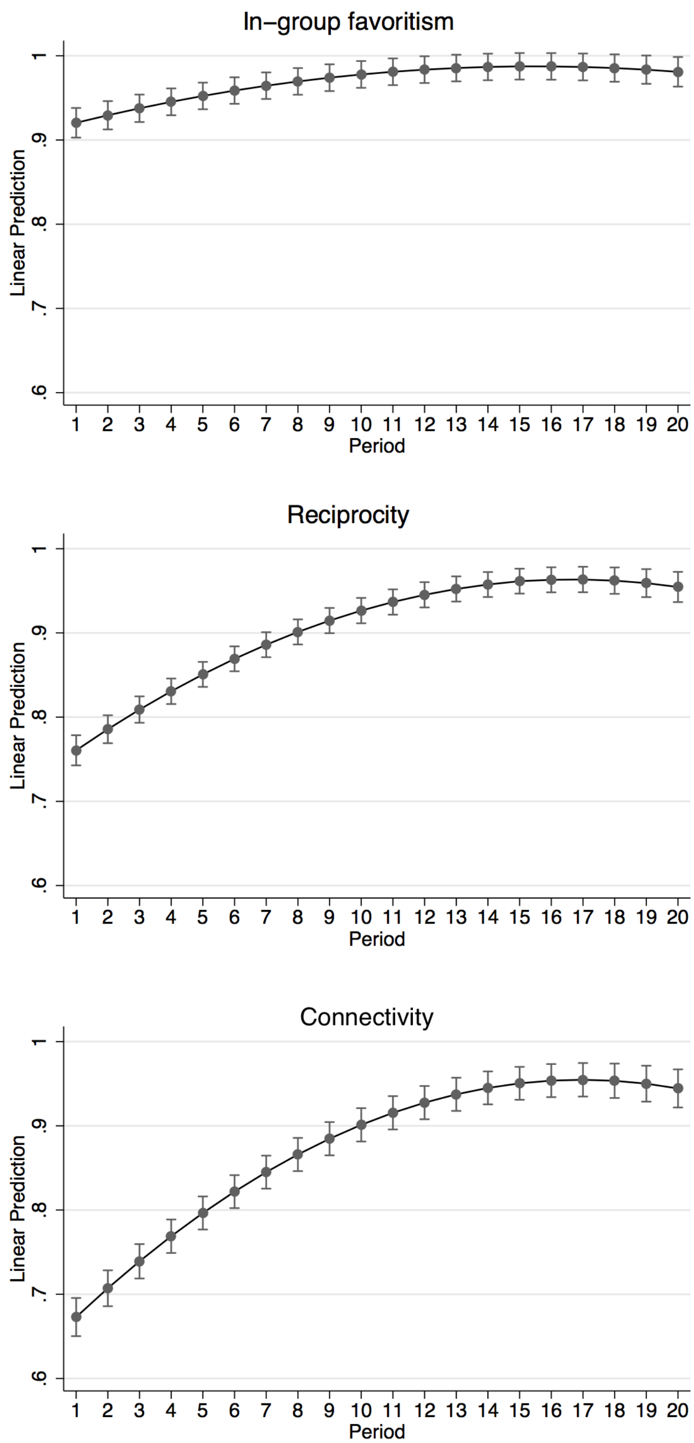


Figure 1. Predictive margins by period.

Majority groups find it harder to reach affiliation consensus, which is not the case when all subjects have aligned preferences in the population.

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 they make.

Our model indicated that the choice an actor makes about what behaviour to adopt depends on her identity and the influence of others around her. An actor wants to coordinate with the highest number of neighbours 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.

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 behaviour 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 behaviour they prefer given their identity. This first result reinforces the categorisation argument of identity theory showing how identities can be so strong that are used to help focalise 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. That is, actors aim to connect completely within their component. The conflict in preferences makes the payoff dominant 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 [27], 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 (i.e. in-group).

Our third empirical result is a surprising observation. When there are conflicting preferences and individuals segregate favouring only their in-groups being in the minority facilitates coordination and stability. Thus, the minority groups tended to the complete connection 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 maximising 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 (e.g. minimal group paradigm). In-group bias is significantly observed, even for cases where there is a majority and a minority [29]. Our results show 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.

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 behaviour [1], we model social categories as fixed. Our aim was to understand the adoption of behaviour 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 [15]. 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 maximise payoffs by integrating with their out-group. On the other, communication could reinforce in-group bias and segregation, by facilitating higher levels of connectivity between players sharing common identities.

Appendix A. Subjects' instructions.

LABORATORIO DE INVESTIGACIÓN EN ECONOMÍA EXPERIMENTAL

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 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 symbo 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 <u>the same action you choose</u>. 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.

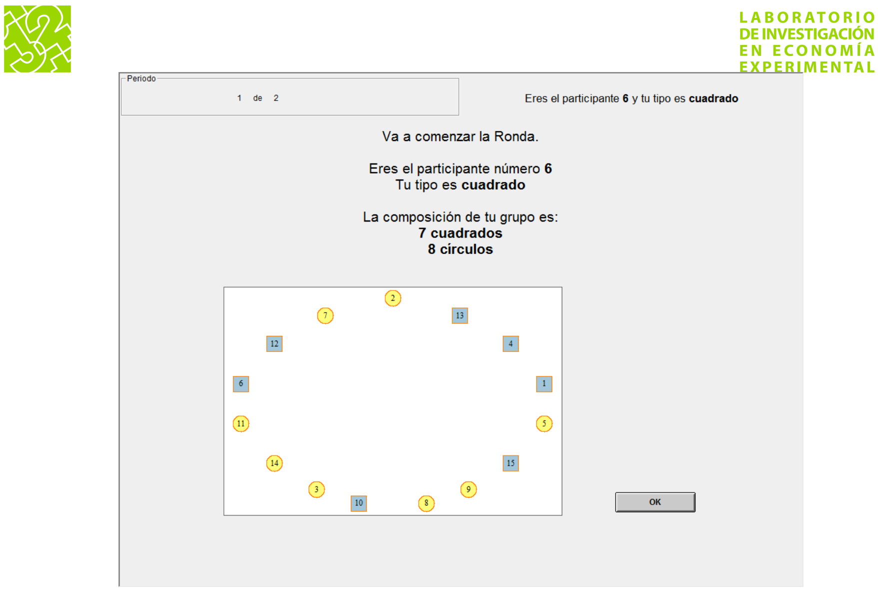
You are ***square***:
- If you choose ***down*** you get **6 points for each** coordination with your connections.
- If you choose ***up*** you get **4 points for each** coordination with your connections.

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):

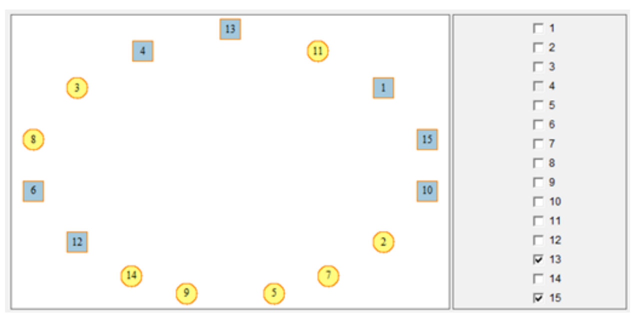
Begining of a round:

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.

[***<u>Text in the image</u>***: 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**.]



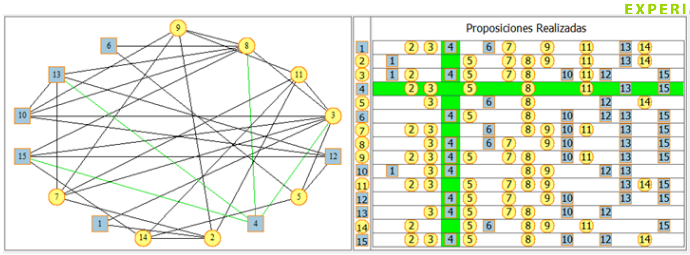
Part 1- Proposals:



The first decisión 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:

LABORATORIO DE INVESTIGACIÓN EN ECONOMÍA EXPERIMENTAL



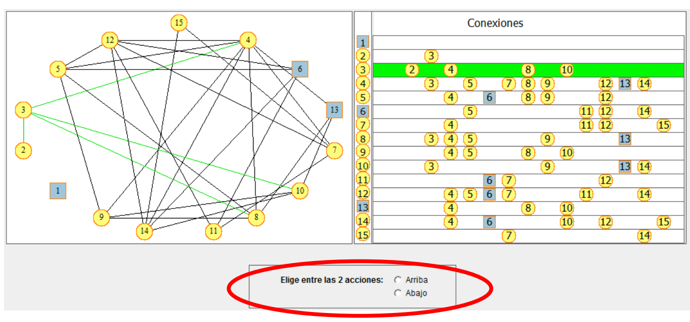
Once all the participants have proposed the connections they want to form you will see the network of connections that results. A connection is formed when two participants propose to each other. In the picture of the network you will see your connections highlighted in green, and on the table next to 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 conenctions 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**:

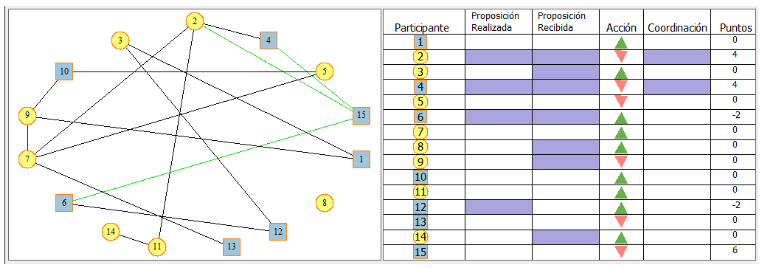
- 6 points for each coordination in ↑
- 4 points for each coordination in ↓

LABORATORIO
DE INVESTIGACIÓN
EN ECONOMÍA
EXPERIMENTAL

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):

LABORATORIO
DE INVESTIGACIÓN
EN ECONOMÍA
EXPERIMENTAL

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

Acknowledgments. The authors declare that the funding for this project was provided by the Spanish Ministry of Science and Technology and the European Feder Founds for financial support under project ECO2013-46550-R and Generalitat Valenciana PROMETEOII/2014/054.

REFERENCES

[1] G. A. Akerlof and R. E. Kranton, Economics and identity, *The Quarterly Journal of Economics*, **115** (2000), 715–753.

[2] G. A. Akerlof and R. E. Kranton, Identity and schooling: Some lessons for the economics of education, *Journal of economic literature*, **40** (2002), 1167–1201.

[3] G. A. Akerlof and R. E. Kranton, Identity and the economics of organizations, *The Journal of Economic Perspectives*, **19** (2005), 9–32.

[4] G. A. Akerlof and R. E. Kranton, *Identity Economics: How Our Identities Shape Our Work, Wages, and Well-Being*, Princeton University Press, 2010.

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

[6] C. Ballester, A. Calvó-Armengol and Y. Zenou, Who's who in networks. wanted: The key player, *Econometrica*, **74** (2006), 1403–1417.

[7] H. Bernhard, E. Fehr and U. Fischbacher, Group affiliation and altruistic norm enforcement, *The American Economic Review*, **96** (2006), 217–221.

[8] M. Billig and H. Tajfel, Social categorization and similarity in intergroup behaviour, *European Journal of Social Psychology*, **3** (1973), 27–52.

[9] M. B. Brewer, In-group bias in the minimal intergroup situation: A cognitive-motivational analysis, *Psychological bulletin*, **86** (1979), 307–324.

[10] A. Calvó-Armengol and R. İlkılıç, Pairwise-stability and nash equilibria in network formation, *International Journal of Game Theory*, **38** (2009), 51–79.

[11] C. Camerer, *Behavioral Game Theory: Experiments in Strategic Interaction*, Princeton University Press, 2003.

[12] G. Charness, L. Rigotti and A. Rustichini, Individual behavior and group membership, *The American Economic Review*, 1340–1352.

[13] Y. Chen and S. X. Li, Group identity and social preferences, *The American Economic Review*, **99** (2009), 431–457.

[14] G. Ellison, Learning, local interaction, and coordination, *Econometrica*, **61** (1993), 1047–1071.

[15] J. Farrell and M. Rabin, Cheap talk, *Journal of Economic Perspectives*, **10** (1996), 103–118.

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

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

[18] L. Goette, D. Huffman and S. Meier, The impact of group membership on cooperation and norm enforcement: Evidence using random assignment to real social groups, *The American economic review*, **96** (2006), 212–216.

[19] S. Goyal, *Connections: An Introduction to the Economics of Networks*, Princeton University Press, 2007.

[20] M. Granovetter, Threshold models of collective behavior, *American journal of sociology*, **83** (1978), 1420–1443.

[21] P. Hernández, M. Muñoz-Herrera and Á. Sánchez, Heterogeneous network games, *Games and Economic Behavior*, **79** (2013), p56.

[22] M. O. Jackson, A survey of network formation models: Stability and efficiency, *Group Formation in Economics: Networks, Clubs, and Coalitions*, (2010), 11–57.

[23] M. O. Jackson, Networks and economic behavior, *Annu. Rev. Econ.*, **1** (2009), 489–511.

[24] M. O. Jackson, *Social and Economic Networks*, Princeton University Press, 2008.

[25] M. O. Jackson and A. Watts, The existence of pairwise stable networks, Seoul Journal of Economics.

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

[27] M. O. Jackson and A. Wolinsky, A strategic model of social and economic networks, *Journal of economic theory*, **71** (1996), 44–74.

[28] M. Kandori, G. J. Mailath and R. Rob, Learning, mutation, and long run equilibria in games, *Econometrica: Journal of the Econometric Society*, **61** (1993), 29–56.

[29] G. J. Leonardelli and M. B. Brewer, Minority and majority discrimination: When and why, *Journal of Experimental Social Psychology*, **37** (2001), 468–485.

[30] D. López-Pintado, Contagion and coordination in random networks, *International Journal of Game Theory*, **34** (2006), 371–381.

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

[32] M. McPherson, L. Smith-Lovin and J. M. Cook, Birds of a feather: Homophily in social networks, *Annual review of sociology*, **27** (2001), 415–444.

[33] S. Morris, Contagion, *The Review of Economic Studies*, **67** (2000), 57–78.

[34] B. Mullen, R. Brown and C. Smith, Ingroup bias as a function of salience, relevance, and status: An integration, *European Journal of Social Psychology*, **22** (1992), 103–122.

[35] M. Muñoz-Herrera, J. Dijkstra, A. Flache and R. Wittek, 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.

[36] T. Schelling, *Micromotives and Macrobehavior*, New York: WW Norton, 1978.

[37] T. Snijders and R. J. Bosker, *Multilevel Analysis: An Introduction to Basic and Applied Multilevel Analysis*, 2nd edition, London: Sage, 2012.

[38] H. Tajfel and J. C. Turner, An integrative theory of intergroup conflict, *The social psychology of intergroup relations*, **33** (1979), p47.

[39] J. C. Turner, Social categorization and social discrimination in the minimal group paradigm, *Differentiation between social groups: Studies in the social psychology of intergroup relations*, 101–140.

[40] J. C. Turner, M. A. Hogg, P. J. Oakes, S. D. Reicher and M. S. Wetherell, *Rediscovering the Social Group: A Self-Categorization Theory.*, Basil Blackwell, 1987.

[41] F. Vega-Redondo, *Complex Social Networks*, vol. 44, Cambridge Univ Pr, 2007.

[42] L. M. Verbrugge, The structure of adult friendship choices, *Social Forces*, **56** (1977), 576–597.

[43] X. Vives, Games with strategic complementarities: New applications to industrial organization, *International Journal of Industrial Organization*, **23** (2005), 625–637.

[44] H. P. Young, The evolution of conventions, *Econometrica: Journal of the Econometric Society*, **61** (1993), 57–84.

Received February 2016; revised April 2016.

E-mail address: penelope.hernandez@uv.es
E-mail address: ellwardt@wiso.uni-koeln.de
E-mail address: guillem.martinez@uv.es
E-mail address: edson.munoz@uv.es