A belief-based theory of homophily[*]

Willemien Kets[†] Alvaro Sandroni[‡]

September 29, 2016

Abstract

We introduce a model of homophily that does not rely on the assumption of homophilous preferences. Rather, it builds on the dual process account of Theory of Mind in psychology which focuses on the role of introspection in decision making. Homophily emerges because players find it easier to put themselves into the shoes of a member of their own group. Endogenizing the drivers of homophily permits us to derive novel comparative statics results and to explain commonly observed features of social and economic networks.

[*] We thank Sandeep Baliga, Vincent Crawford, Georgy Egorov, Tim Feddersen, Matthew Jackson, Rachel Kranton, Yuval Salant, Paola Sapienza, Rajiv Sethi, Eran Shmaya, Andy Skrzypacz, Jakub Steiner, Jeroen Swinkels, and numerous seminar audiences and conference participants for helpful comments and stimulating discussions. Alex Limonov provided research assistance.

[†] Kellogg School of Management, Northwestern University. E-mail: w-kets@kellogg.northwestern.edu

[‡] Kellogg School of Management, Northwestern University. E-mail: sandroni@kellogg.northwestern.edu

1. Introduction

Homophily, the tendency of people to interact with similar people, is a widespread phenomenon that has been studied in a variety of different fields, ranging from economics (Currarini, Jackson, and Pin, 2009), to organizational research (Borgatti and Foster, 2003), social psychology (Gruenfeld and Tiedens, 2010), political science (Mutz, 2002), and sociology (McPherson, Smith-Lovin, and Cook, 2001). Homophily can give rise to segregated social and professional networks, and can affect investment in education (Calvó-Armengol, Patacchini, and Zenou, 2009), wages and employment (Patacchini and Zenou, 2012), and diffusion of information (Golub and Jackson, 2012). Thus, understanding the root sources of homophily is of paramount importance.

Much of the existing literature explains homophily by assuming a direct preference for associating with similar others (see Jackson, 2014, for a survey). However, without a theory of the determinants of these preferences, it is hard to explain why homophily is observed in some cases, but not in others (beyond positing homophilous preferences only in the former settings).

We provide a theory of homophily that does not assume homophilous preferences. Rather, in our model, a preference to interact with similar others is a natural outcome of individuals' desire to reduce strategic uncertainty. This framework delivers new testable implications and allows us to explain commonly observed features of social and economic networks.

In our model, players belong to different groups that differ in their mental models, i.e., perspectives, interpretations, narratives, and worldviews (Craik, 1943). Shared mental models facilitate social interactions. Many social interactions are governed by unwritten rules or customs that prescribe the appropriate course of action.[1] These prescriptions are not universal. In particular, the prescription may be sensitive to the context. For example, offering to share food may be appropriate in some settings (e.g., family dinners) and not in others (e.g., business dinners), while in others it is unclear (e.g., social outings with colleagues). Individuals that share the same mental model tend to view the same prescription as focal, while individuals with different mental models are more likely to disagree on what is focal (e.g., is this situation more like a formal dinner or a family-style meal?); see Denzau and North (1994). Thus, the likelihood that individuals follow the same prescription is greater when they belong to the same group.[2] Accordingly, if individuals benefit from coordinating their activities, then they

[1] These ideas have a long history in both philosophy (Hume (1740), Lewis, 1969) and in economics (Schelling, 1960). See, e.g., Camerer and Vepsailanen (1988), Bacharach (1993), Sugden (1995), Skyrms (1996), Sugden (1998), Camerer and Knez (2002), Weber and Camerer (2003), and Camerer and Weber (2013) for more recent perspectives.

[2] This is in line with the suggestion of Kreps (1990) that culture is a source of focal principles. There is

have an incentive to associate with members of their own group.

While intuitive, it is difficult to capture these ideas using standard models. Standard game-theoretic models cannot model how mental models affect behavior. Standard models are thus unable to explain why players may find it easier to coordinate with their own group. To formally model these ideas, we build on the dual process account of Theory of Mind in psychology (see Kets and Sandroni, 2015). The dual process account of Theory of Mind is an influential theory in psychology that posits that an individual initially reacts instinctively, and then adapt his views by reasoning about what he would do if he were in the opponent's position.[3]

To capture this, we assume that each player has some initial (random) impulse telling him which action is appropriate. A player's first instinct is to follow his impulse. By introspection, the player realizes that the opponent may also have an impulse to choose a certain action. In addition, he realizes that if his opponent is similar to him, his opponent's impulse may be similar to his own. So, impulses can be used to form initial beliefs. This may lead the player to adjust his action and not act on impulse. The reasoning does not stop here, however. If the player thinks a little more, he realizes that his opponent may have gone through a similar reasoning process and may have adjusted her action. This leads the player to revise his initial belief, and so on, up to arbitrarily high order. The limit of this procedure, where players go through the entire reasoning process in their mind before taking a decision, defines an *introspective equilibrium.*

In our model, players are matched to play a coordination game. In line with evidence from neuroscience and psychology (Elfenbein and Ambady, 2002; de Vignemont and Singer, 2006), we assume that players find it easier to anticipate the instinctive reactions of members of their own group. Thus, impulses are more strongly correlated within groups than across groups.

When players use introspection to decide on their action, they are more successful at coordinating with their own group, as we show. As a result, they face less *strategic uncertainty* when interacting with players who are similar to them. This gives players an incentive to associate with their own group. We consider an extended game where players can seek out members of their own group by choosing the same *project* (e.g., a hobby, profession, or neighborhood). The resulting level of homophily can be high even if it is costly to seek out the

substantial experimental evidence that there is more coordination and less strategic uncertainty when players interact with their own group; see Section 1.1.

[3] See Epley and Waytz (2010) for a survey of the research on Theory of Mind in psychology. The dual process account of Theory of Mind relies on a rapid instinctive process and a slower cognitive process. As such, it is related to the two-systems account of decision-making under uncertainty, popularized by Kahneman (2011). The foundations of dual processes theory go back to the work of the psychologist William James (1890/1983).

own group and players do not have a direct preference for interacting with their own group. In this sense, introspection and players' desire to reduce strategic uncertainty are root causes of homophily.

Our model produces unambiguous and intuitive comparative statics. In our model, the level of homophily is determined uniquely and it depends on economic incentives and the similarity in impulses. The level of homophily is higher if the stakes are high and if group members are more similar (i.e., impulses are strongly correlated within a group). In fact, regardless of the distribution over impulses, when coordination payoffs are high, the level of homophily is necessarily high and is above and beyond what can be expected based on direct preferences over projects or groups. Thus, the model produces a series of novel testable implications. These predictions are difficult to obtain with standard game-theoretic models or common refinements, as these do not incorporate the effect of identity on reasoning.

In an extension of the model, players choose how much (costly) effort to invest in meeting others. The level of homophily can now be even higher. Intuitively, there can be a feedback effect: if one group is the dominant group for a given project, in the sense that the majority of the project participants belong to this group, members of this group have a greater incentive to invest effort, even if the majority is only slightly larger than the minority. This, in turn, increases the chances that members of the dominant group are matched with their own group, further enhancing their incentives to form connections. This may lead more players from the dominant group to choose the project, leading the majority to grow.

The cost of forming connections determines whether small differences in initial conditions are amplified. If the effort cost is low, there is a substantial incentive to form connections. This enhances the incentives to segregate, further increasing the incentives for the dominant group to network. The resulting network consists of a tightly connected core of players from the dominant group with a periphery of loosely connected members of the minority group, in line with empirical observations (Jackson, 2014). On the other hand, if the effort cost is high, the incentives to network are attenuated even for the dominant group. This reduces the incentives to segregate, which further damps the incentives to form connections. The result is a sparse network with low levels of homophily. This provides novel testable hypotheses about how properties of networks change when the fundamentals vary. For example, the theory predicts that high levels of homophily go hand in hand with a core-periphery structure.

This paper is organized as follows: after a brief literature review in Section 1.1, we present our basic model in Section 2. Section 3 characterizes the level of homophily in the benchmark model and presents the comparative statics. Section 4 extends the model to allow players to choose their effort and uses this to study network formation. Section 5 concludes. Appendix B shows that the main insights extend to settings where players signal their identity. All proofs can be found in the appendices.

1.1. Related literature

The literature on homophily typically assumes homophilous preferences and investigates the implications for network structure and economic outcomes (e.g., Currarini, Jackson, and Pin, 2009; Bramoullé, Currarini, Jackson, Pin, and Rogers, 2012; Golub and Jackson, 2012; Alger and Weibull, 2013), with Baccara and Yariv (2013, 2016) and Pęski (2008) being notable exceptions. In a public good provision model, Baccara and Yariv show that groups are stable only if their members have similar preferences. Pęski shows that segregation is possible if players have preferences over the interactions that their opponents have with other players.[4] We propose a novel mechanism through which homophily can arise: players have an incentive to interact with similar others if that reduces strategic uncertainty. This allows us to explain the emergence of *value homophily*, that is, homophily based on similarities in attitudes and beliefs (McPherson, Smith-Lovin, and Cook, 2001).

Mental models are often associated with identity or culture. Following the seminal work of Akerlof and Kranton (2000), an emerging literature in economics studies the effect of identity on economic outcomes. In much of this literature, an agent's identity affects his payoffs, not his reasoning. Our emphasis on the cognitive aspects of identity is consistent with a large and growing literature in sociology and anthropology that views identity and culture as being composed of mental models (or "schemas") to interpret the world; see, e.g., DiMaggio (1997). By modeling a player's identity in terms of mental models, we are able to address a novel set of questions such as how homophily varies with economic incentives.[5] Greif (1994) stresses that cultural beliefs may have an important impact on economic outcomes, but notes that a formal analysis of the relations between cultural beliefs and economic outcomes is challenging, because of the multiplicity that can result when beliefs are unrestricted. Here we impose simple and intuitive assumptions on beliefs that are in line with experimental evidence which allow us to derive unique predictions in a range of settings. Kuran and Sandholm (2008) take the culture of a group to be defined by the preferences and equilibrium behaviors of its members. In our model, groups may differ in their equilibrium behavior even if they have identical preferences. Van den Steen (2010) shows that shared beliefs may lead to faster coordination in a learning context, but does not study homophily.[6] Crémer (1993) define culture as the shared knowledge base of a group, and Kreps (1990) suggests that culture is a source of focal principles that

[4] Also see Pęski and Szentes (2013).

[5] In Kets and Sandroni (2015), we explore the conditions under which diversity is optimal when players' identity affects their strategic reasoning.

[6] Also, Van den Steen's (2010) result requires payoff heterogeneity. No such heterogeneity is needed here.

can help select an equilibrium. In our model, players who share a similar background have similar behavioral tendencies. This provides a formal mechanism through which culture can aid equilibrium selection.

The process we consider bears some resemblance with level-k models (Nagel, 1995; Stahl and Wilson, 1995; Costa-Gomes, Crawford, and Broseta, 2001; Costa-Gomes and Crawford, 2006; Crawford, Costa-Gomes, and Iriberri, 2013).[7] There are two important differences. First, while the level-k literature focuses on deviations from equilibrium, we use the reasoning process to select a unique equilibrium. Second, introducing payoff-irrelevant impulses into the model permits us to study the effects of identity on reasoning.

Modeling the introspective process allows us to select a unique outcome in a range of games. This allows us to derive clear comparative statics. This is not possible using a standard equilibrium analysis (see Appendix A). Like other models that are used to explain homophily and segregation, the games we study have multiple equilibria with sometimes very different properties. Other papers have dealt with equilibrium multiplicity by focusing on the subset of equilibria that satisfy a stability property (e.g., Alesina and La Ferrara, 2000; Bénabou, 1993; Sethi and Somanathan, 2004). However, such refinements are not always strong enough to give uniqueness. In particular, in our setting, standard refinements have no bite. We thus need a novel approach to obtain a unique prediction. As we show, taking into account players' reasoning process is a powerful method to obtain uniqueness in a range of different settings.

Our work sheds light on experimental findings that social norms and group identity can help players coordinate effectively, as in the minimum-effort game (Weber, 2006; Chen and Chen, 2011), communication tasks (Weber and Camerer, 2003), the provision point mechanism (Croson, Marks, and Snyder, 2008), risky coordination games (Le Coq, Tremewan, and Wagner, 2015), and Battle of the Sexes (Charness, Rigotti, and Rustichini, 2007; Jackson and Xing, 2014). Chen and Chen (2011) explain the high coordination rates on the efficient equilibrium in risky coordination games in terms of social preferences. Our model provides an alternative explanation, based on beliefs: players are better at predicting the actions of players who belong to the same group. Our mechanism operates even if no equilibrium is superior to another, as in pure coordination games.

[7] A closely related model is the cognitive hierarchy model (Camerer, Ho, and Chong, 2004). This model is less closely related to ours; see Kets and Sandroni (2015).

2. Coordination and introspection

There are two groups, A and B, each consisting of a unit mass of players. Members of these groups are called A-players and B-players, respectively. Group membership is unobservable.[8]

Players are matched in pairs. Each player is matched with a member of his own group with probability $\hat{p} \in (0,1]$. In this section, the probability $\hat{p}$ is exogenous. In Section 3, we endogenize $\hat{p}$. Players who are matched play a coordination game, with payoffs given by:

$$\begin{array}{c|c|c|} & s^1 & s^2 \\ \hline s^1 & v,v & 0,0 \\ \hline s^2 & 0,0 & v,v \\ \hline \end{array} \;,\; v > 0.$$

Payoffs are commonly known. The precise assumptions on payoffs are not critical: the key assumption is that players have an incentive to coordinate.[9]

The game has two strict Nash equilibria: one in which both players choose s^1, and one in which both players choose s^2. Thus, players cannot deduce from the payoffs alone how others will behave. So, there is significant *strategic uncertainty*: players do not know what the opponent will do. We build on the dual process account of Theory of Mind in psychology to describe how players model their opponent and choose their actions. According to the dual process account, people have impulses, and through introspection (i.e., by observing their own impulse) players can learn about the impulses of others and thus form a conjecture about their opponent's behavior. Given this newly-formed conjecture, it may be optimal for a player to deviate from his initial impulse. Upon introspection, they may realize that their opponent may likewise adjust their behavior. In turn, this may lead them to revise their initial conjecture, and so on (Kets and Sandroni, 2015).[10]

This is formalized as follows. Each player j receives an *impulse* $i_j = 1, 2$. Impulses are payoff-irrelevant, privately observed signals. If a player's impulse is 1, then his initial impulse

[8] In our model, groups differ in the mental models that their members hold. While mental models are in principle unobservable, there may be observable proxies. For example, in some cases, different demographic groups tend to use different mental models (see, e.g., Page, 2007, for a discussion). If that is the case, group membership (i.e., which mental model a player uses) is imperfectly observable. Our results extend to this setting.

[9] For example, our results go through if payoffs are asymmetric, if one Nash equilibrium Pareto-dominates the other, or if one of the Nash equilibria is riskier than the other, as long as the game is a low-potential game in the terminology of Kets and Sandroni (2015). The results also extend to settings where there are skill complementarities across groups, as long as they are not too strong.

[10] Robalino and Robson (2015) interpret Theory of Mind as the ability to learn other players' payoffs, and shows that this confers an evolutionary benefit in volatile environments.

is to take action s^1. Likewise, if a player's impulse is 2, then his initial impulse is to choose action s^2.

Impulses may be partly shaped by players' mental models. People with shared mental models will often respond in a similar way to a given strategic situation, while people with different mental models may respond differently. Players thus find it easier to anticipate the instinctive response of members of their own group. To model this, we assume that impulses are more strongly correlated within groups than across groups. Specifically, each group $g = A, B$ is characterized by a state $\theta_g = 1, 2$. A priori, θ_g is equally likely to be 1 or 2; and the states of different groups are independent. Conditional on $\theta_g = m$, a g-player has an initial impulse to choose action s^m with probability $q \in (\frac{1}{2}, 1)$, independently across players.

This simple model ensures that players are more likely to have the same impulse if they belong to the same group. The *within-group similarity* is the probability $Q_{in} = Q_{in}(q)$ that two group members receive the same impulse. In the appendix, we show that Q_{in} is strictly between $\frac{1}{2}$ and 1, while the probability Q_{out} that two players from different groups have the same impulses is $Q_{out} = \frac{1}{2}$. In words, a player's impulse is more informative of the impulse of a member of his own group than of a player outside the group.

A player's first instinct is to follow his initial impulse, without any strategic considerations. This defines the level-0 strategy σ_j^0 for player j. Through introspection, a player realizes his opponent likewise follows his impulse. By observing his own impulse, a player can form a belief about his opponent's impulse and formulate a best response against the belief that the opponent follows her impulse. This defines the player's level-1 strategy σ_j^1. In general, at level $k > 1$, a player formulates a best response against his opponent's level-$(k-1)$ strategy. This, in turn, defines his level-k strategy σ_j^k. Together, this defines a reasoning process with infinitely many levels. The levels are merely constructs in a player's mind. We are interested in the limit of this process as the level k goes to infinity. If there is a limiting strategy σ_j for each player j, then the profile $\sigma = (\sigma_j)_j$ is an *introspective equilibrium.*

In an introspective equilibrium, group identity influences behavior because it affects impulses and beliefs. This is true even if groups are identical in terms of payoffs, as we assume here.

Every introspective equilibrium is a correlated equilibrium, so that behavior in an introspective equilibrium is always consistent with common knowledge of rationality (Aumann, 1987):

Proposition 2.1. [Rationality of Introspective Equilibrium, Kets and Sandroni, 2015] *Every introspective equilibrium is a correlated equilibrium.*

So, while introspective equilibrium is based on ideas from psychology and assumes that players' initial reaction is nonstrategic, it does not presume that players are boundedly rational.

Perhaps surprisingly, instinctive reactions can in fact be consistent with equilibrium: the seemingly naive strategy of following one's initial impulse is the optimal strategy that results from the infinite process of high-order reasoning, as the next result shows.

Proposition 2.2. [Introspective Equilibrium Coordination Game, Kets and Sandroni, 2015] *There is a unique introspective equilibrium of the coordination game. In this equilibrium, each player follows his initial impulse.*

So, in this case, the reasoning process delivers a simple answer: it is optimal to act on instinct. Intuitively, the initial appeal of following one's impulse is reinforced at higher levels, through introspection: if a player realizes that his opponent follows her impulse, it is optimal for him to do so as well; this, in turn, makes it optimal for the opponent to follow her impulse.

Proposition 2.2 shows that introspection allows players to coordinate. However, coordination is imperfect: by Lemma C.1 in the appendix, the coordination rate lies strictly between $\frac{1}{2}$ and 1. So, while introspective equilibrium is consistent with common knowledge of rationality, it predicts non-Nash behavior in this environment. Such non-Nash behavior is in fact observed experimentally. Mehta, Starmer, and Sugden (1994) show that while subjects are unable to coordinate on one of the pure-strategy Nash equilibria of a coordination game, they tend to coordinate at a higher rate than in the mixed equilibrium. Also, subjects are more likely to coordinate when one of the alternatives is highly focal for a group (Bardsley, Mehta, Starmer, and Sugden, 2009). In line with this observation, our model predicts that the coordination rate is high if players' impulses tend to agree (i.e., Q_{in} close to 1).

These predictions are intuitive, but cannot be obtained using standard methods as they require a way to model how mental models shape strategic reasoning. In addition to providing intuitive predictions, the introspective process also selects a unique outcome, even in a setting with many (correlated) equilibria where standard equilibrium refinements have no bite. The uniqueness of introspective equilibrium will be critical for deriving unambiguous comparative static results in Sections 3 and 4.

In the unique introspective equilibrium, expected payoffs are:

$$\left[\hat{p}Q_{in} + (1-\hat{p}) \cdot Q_{out}\right] \cdot v.$$

The *marginal benefit of interacting with the own group* is the change in payoffs when the probability of interacting with the own group increases. Thus, it is defined by

$$\beta := (Q_{in} - Q_{out}) \cdot v.$$

Since $Q_{in} > Q_{out}$, the marginal benefit of interacting with the own group is strictly positive. Thus:

Corollary 2.3. *The expected utility of a player strictly increases with the probability $\hat{p}$ of being matched with a player from the own group.*

This follows because players are more likely to coordinate with members of their own group, consistent with experimental evidence (see Section 1.1). Thus, a player's expected payoff increases with the probability that he interacts with his own group. This means that players have an incentive to seek out similar players, that is, to be homophilous. We explore the implications in the next section.

3. Homophily

In ordinary life, there is often no exogenous matching mechanism. People meet after they have independently chosen a common place or a common activity. Accordingly, we model an extended game in which there are two *projects* (e.g., occupations, clubs, neighborhoods), labeled a and b. Players first choose a project and are then matched uniformly at random with someone that has chosen the same project. Once matched, players play the coordination game described in Section 2.

Each player has an intrinsic value for each project. Players in group A have a slight tendency (on average) to prefer project a. Specifically, for each A-player j, the value $w_j^{A,a}$ of project a is drawn uniformly at random from $[0, 1]$, while the value $w_j^{A,b}$ of project b is drawn uniformly at random from $[0, 1-2\varepsilon]$, for some small $\varepsilon > 0$. For B-players, an analogous statement holds with the roles of projects a and b reversed. So, on average, B-players have a slight tendency to prefer project b. Values are drawn independently (across players, projects, and groups). Under these assumptions, a fraction $\frac{1}{2}+\varepsilon$ of A-players intrinsically prefer project a, and a fraction $\frac{1}{2}+\varepsilon$ of B-players intrinsically prefers project b; see Appendix C.2 for details. Thus, project a is the *group-preferred project* for group A, and project b is the group-preferred project for group B.[11]

Players' payoffs are the sum of the intrinsic value of the chosen project and the (expected) payoff in the unique introspective equilibrium of the coordination game. By Proposition 2.2, the expected payoff of an A-player with project a is thus

$$v \cdot \left[\hat{p}_A \cdot Q_{in} + (1-\hat{p}_A) \cdot Q_{out}\right] + w_j^{A,a},$$

for any given probability $\hat{p}_A$ of interacting with the own group; and likewise for other projects and groups.

[11] It is not critical for our results that groups differ in their preferences over projects. What we need is that groups differ in their instinctive choice of project (on average).

To choose their project, players follow the same introspective process as before, taking into account their payoffs in the coordination game in the second stage. At level 0, players follow their impulse and select the project they intrinsically prefer. At level $k > 0$, players formulate a best response to the project choices at level $k - 1$, given their intrinsic preferences. That is, a player chooses project a if and only if the expected payoff from project a is at least as high as from b, given the choices at level $k - 1$. Let p_k^a be the fraction of A-players among those with project a at level k, and let p_k^b be the fraction of B-players among those with project b at level k. The limiting behavior, as k increases, is well-defined.

Lemma 3.1. [Convergence of Introspective Process] *The limit p^π of the fractions $p_0^\pi, p_1^\pi, \ldots$ exists for each project $\pi = a, b$. Moreover, the limits are the same for both projects: $p^a = p^b$.*

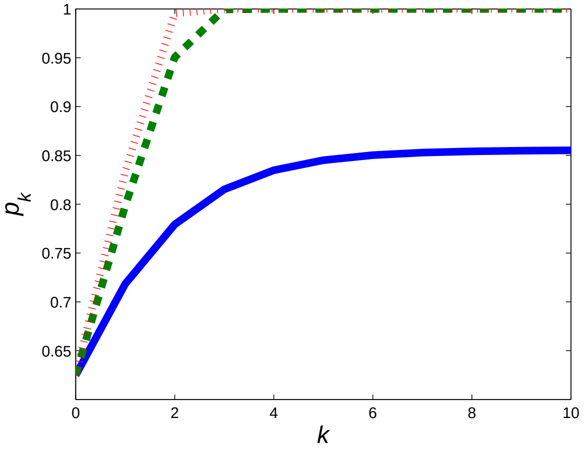

Figure 1: The probability p_k that a player chooses the group-preferred project at level k, for $\beta = 0.4$ (solid line), $\beta = 0.8$ (dashed line), and $\beta = 1$ (dotted line), for $\varepsilon > 0$ small.

Figure 1 illustrates the convergence of the introspective process. Let $p := p^a = p^b$ be the limiting probability in the introspective equilibrium. So, p is the probability that a player with the group-preferred project is matched with a player from the same group. Define the *level of homophily* $h := p - \frac{1}{2}$ to be the difference between the probability that a player with the group-preferred project meets a player from the same group in the introspective equilibrium and the probability that he is matched with a player from the same group uniformly at random, independent of project choice. If the level of homophily is close to 0, there is almost full integration. If the level of homophily is close to $\frac{1}{2}$, there is nearly complete segregation.

Since there is only a slight asymmetry in preferences, the initial level of homophily (i.e., the level of homophily based on intrinsic preferences) is minimal: $h^0 := \varepsilon$. However, as the next result shows, the equilibrium level of homophily can be high:

Proposition 3.2. [Homophily: Equilibrium] *There is a unique introspective equilibrium of the extended game. In the unique equilibrium, players follow their impulse in the coordination game, and players' project choices give rise to complete segregation ($h = \frac{1}{2}$) if and only if*

$$\beta \geq 1 - 2\varepsilon.$$

If segregation is not complete ($h < \frac{1}{2}$), then the equilibrium level of homophily is given by:

$$h = \frac{(1-2\varepsilon)}{4\beta^2} \cdot \left[2\beta - 1 + \sqrt{\frac{4\beta^2}{1-2\varepsilon} - 4\beta + 1}\right],$$

where $\beta = v \cdot (Q_{in} - Q_{out})$ is the marginal benefit of interacting with the own group. In any case, the equilibrium level of homophily exceeds the initial level of homophily (i.e., $h > \varepsilon$).

There can be substantial homophily in the unique introspective equilibrium. In that case, most players choose the group-preferred project, even if they have a strong intrinsic preference for the other project. Interactions may thus be homophilous even if players have no direct preference for interacting with members of their own group. Homophily is not the result of any payoff-relevant differences between groups: groups are almost identical (i.e., ε small), and if homophily were based solely on intrinsic preferences, then homophily would be negligible (i.e., $h^0 = \varepsilon$). Instead, homophily is the result of strategic considerations. Strategic considerations *always* produce more homophily than would follow from differences in intrinsic preferences over projects (i.e., $h > \varepsilon$). Introspection and players' desire to reduce strategic uncertainty are thus root causes of homophily. This is consistent with experimental evidence that shows that subjects are more homophilous if interacting with their own group helps reduce strategic uncertainty (Currarini and Mengel, 2016).

When choosing a project, players do not act on impulse. Instead, introspection leads them to reevaluate their initial impulse. At level 1, player realize that there is a slightly higher chance of interacting with members of their own group if they choose the group-preferred project. As a result, players may select the group-preferred project even if they have a slight intrinsic preference for the other project. At level 2, an even higher fraction of agents may select the group-preferred project because players expect the odds of finding a similar player this way to be even higher than at level 1. So, the attractiveness of the group-preferred project is reinforced throughout the entire reasoning process, as illustrated in Figure 1. As a result, the equilibrium level of homophily strictly exceeds the initial level (i.e., $h > h^0 = \varepsilon$).

One implication of Proposition 3.2 is that people who belong to the same group become similar on other dimensions as well, e.g., by choosing the same hobbies, professions, or clubs as other members of their group, consistent with empirical evidence (Kossinets and Watts, 2009). This differs from peer effects, i.e., the well-known phenomenon that individuals who

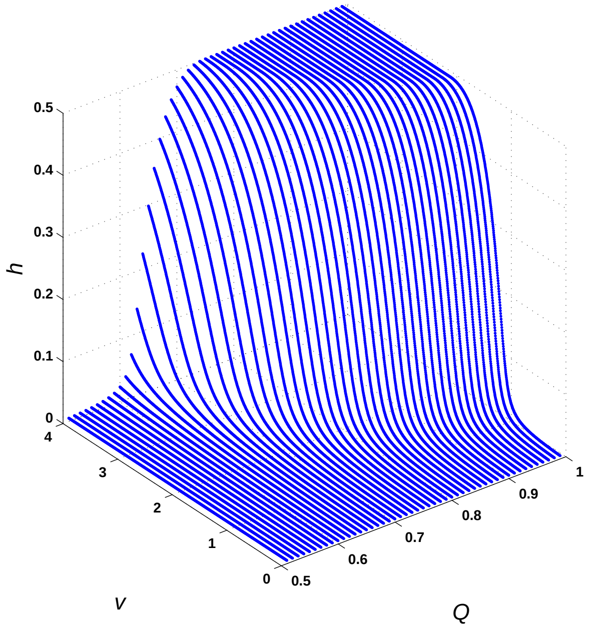

Figure 2: The equilibrium level of homophily h as a function of the coordination payoff v and the within-group similarity Q_{in}.

interact frequently influence each other, and thus become more similar (e.g., Benhabib, Bisin, and Jackson, 2010). Here, being similar is a precondition for interaction, not a result thereof.

The comparative statics for the level of homophily follow directly from Proposition 3.2:

Corollary 3.3. [Homophily: Comparative Statics] *The level of homophily h increases with the within-group similarity Q_{in} and with the coordination payoff v. Cultural factors and economic incentives are complements: homophily is high when either the within-group similarity or the coordination payoff is high.*

Figure 2 shows the level of homophily as a function of the coordination payoff v and the within-group similarity Q_{in}. Regardless of the within-group similarity Q_{in}, the level of homophily increases with economic incentives to coordinate. These comparative statics results deliver clear and testable predictions of the model: there is a positive correlation between coordination payoffs and homophily, regardless of the exact distribution of impulses.

Our results have an interesting implication. Assume that a group, say B, is replaced by another group B', and B'-players are as unpredictable for members of group A as B-players (and vice versa). Then, the model predicts that the level of homophily should remains unchanged. This suggests that identity does not matter per se; what matters is relative predictability.

Corollary 3.3 also demonstrates that groups that have more similar impulses will be more homophilous. This suggests that a similarities can be reinforcing: if group members are similar (i.e., Q_{in} high), then they tend to choose the same project; this, in turn, may lead to more shared experiences and mutual influence, leading them to become even more similar.

These intuitive predictions require some form of equilibrium selection to be properly formalized. We use the dual process account of Theory of Mind to obtain a unique equilibrium. Standard analysis delivers a multiplicity of (correlated) equilibria and cannot deliver unambiguous comparative statics. This is because the set of (correlated) equilibrium changes when parameters are varied; see Appendix A. By contrast, the introspective process selects a unique correlated equilibrium.

Our results do not depend on our specific assumptions, such as the exact assumptions on preferences or the distribution of impulses. For example, the assumption that there are group-preferred projects can be relaxed substantially. All we need is that groups differ in their instinctive reactions when choosing projects. In particular, our results go through if a (large) majority of both groups (intrinsically) prefer a certain project (or have an instinctive impulse to choose that project), as long as there is some asymmetry across groups. Our results also continue to hold if players can "opt out" of the coordination game by choosing an outside option that gives each player a fixed utility independent of which other players choose this option. Finally, the same results obtain in natural variants of the model. Appendix B demonstrates that our results go through if players cannot sort by choosing projects, but instead choose markers, that is, observable attributes such as tattoos or specific attire, to signal their identity and increase the chance of meeting with members of their own group. Again, there can be high levels of homophily in equilibrium.

4. Network formation

In many situations, people can choose the effort spend socializing. So, we extend the basic model to allow players to choose how much (costly) effort they want to invest in meeting others. We show that the basic mechanism that drives homophily may be reinforced and can explain commonly observed properties of social and economic networks.

To analyze this setting, it will be convenient to work with a finite (but large) set of players.[12] Each group $G = A, B$ has N players, so that the total number of players is $2N$. In the first stage, players simultaneously choose effort and projects. This determines the probability that they meet other players. In the second stage, they play the coordination game in Section 2 with the players they met in the first stage.

By investing effort, a player increases the chance that he meets other players. Specifically,

[12] Defining networks with a continuum of players gives rise to technical problems. Our results in Sections 2 and 3 continue to hold under the present formulation of the model (with a finite player set), though the notation becomes more tedious.

if two players j, ℓ have chosen the same project $\pi = a, b$ and invest effort e_j and e_ℓ, respectively, then the probability that they interact is[13]

$$\frac{e_j \cdot e_\ell}{E^\pi},$$

where E^π is the total effort of the players with project π.[14] Thus, efforts are complements: players interact only if they both invest. This is in line with the assumption of bilateral consent in deterministic models of network formation (Jackson and Wolinsky, 1996). By normalizing by the total effort E^π, we ensure that the network does not become arbitrarily dense as the number of players grows large. So, the probability of meeting a member of the own group is endogenous here, as in Section 3. Matching probabilities are now affected not only by players' project choice, as in Section 3, but also by their effort levels.

Players play the coordination game with every player they meet. So, investing effort in the first stage increases their expected payoff in the second stage. However, effort is costly: a player that invests effort e pays a cost $ce^2/2$.

As before, at level 0 players choose the project that they intrinsically prefer. So, the share of players that select the group-preferred project at level 0 is $\frac{1}{2} + \varepsilon$. In addition, each player chooses some default effort $e_0 > 0$, independent of his project or group. At higher levels k, each player formulates a best response to their partners choices at level $k - 1$. As before, each player receives a (single) signal that tells him which action is appropriate in the coordination game. He then plays the coordination game with each of the players he is matched to.[15]

The limiting behavior is well-defined and is independent of the level-0 effort choice:

Lemma 4.1. [Convergence of Introspective Process (Networks)] *The limiting probability p and the limiting effort choices exist and do not depend on the effort choice at level 0 (i.e., e_0).*

As before, we have a unique introspective equilibrium, with potentially high levels of homophily:

Proposition 4.2. [Equilibrium Characterization Networks] *There is a unique introspective equilibrium of the network formation game. In the unique equilibrium, players follow*

[13] See, e.g., Cabrales, Calvó-Armengol, and Zenou (2011) and Galeotti and Merlino (2014) for applications of this model in economics.

[14] To be precise, to get a well-defined probability, if $E^\pi = 0$, we take the probability to be 0; and if $e_j \cdot e_\ell > E^\pi$, we take the probability to be 1.

[15] We allow players to take different actions in each of the (two-player) coordination games he is involved in. Nevertheless, in any introspective equilibrium, a player chooses the same action in all his interactions, as it is optimal for him to follow his impulse (Proposition 2.2).

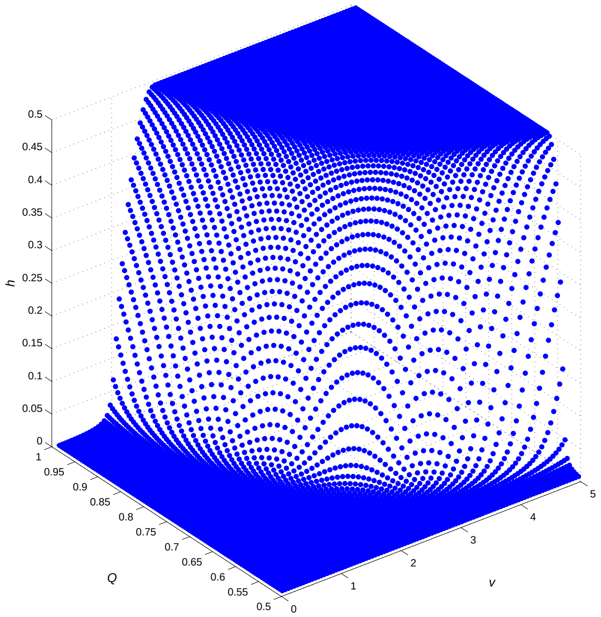

Figure 3: The level of homophily h as a function of the coordination payoff v and the within-group similarity Q_{in} $(c = 1)$.

their impulse in the coordination game, and they choose positive effort in the network formation stage. Players with the group-preferred project exert strictly more effort than players with the other project. In all cases, the fraction of players choosing the group-preferred project exceeds the initial level (i.e., $h > \varepsilon$).

As before, players segregate for strategic reasons and the level of homophily is greater than what would be expected on the basis of intrinsic preferences alone (i.e., $h > \varepsilon$). Importantly, players with the group-preferred project invest more effort in equilibrium than players with the other project. This is intuitive: a player with the group-preferred project has a high chance of meeting people from her own group, and thus a high chance of coordinating successfully. In turn, this reinforces the incentives to segregate.

Figure 3 illustrates the comparative statics of the unique equilibrium. As before, the level of homophily increases with the within-group similarity and with economic incentives, and the two are complements.

While the proof of Proposition 4.2 provides a full characterization of the equilibrium, the comparative statics cannot be analyzed analytically, as the effort levels and the level of homophily depend on each other in intricate ways. We therefore focus on deriving analytical results for the case where the network becomes arbitrarily large (i.e., $|N| \to \infty$). As a first step, we give an explicit characterization of the unique introspective equilibrium:

Proposition 4.3. *Consider the limit where the number of players grows large (i.e., $|N| \to \infty$). The effort chosen by the players with the group-preferred project in the unique introspective equilibrium converges to*

$$e^* = \frac{v}{4c} \cdot \left(1 + 2Q_{in} - \frac{1}{2h} + \sqrt{4Q_{in}^2 - 1 + \frac{1}{4h^2}} \right),$$

while the effort chosen by the players with the other project converges to

$$e^- = \frac{v}{c} \cdot (Q_{in} + \tfrac{1}{2}) - e^*,$$

which is positive but below the effort e^* (i.e., $e^- \in (0, e^*)$).

Proposition 4.3 shows that in the unique introspective equilibrium, the effort levels depend on the level of homophily. The level of homophily, in turn, is a function of the equilibrium effort levels. For example, by increasing her effort, an A-player with the group-preferred project a increases the probability that players from both groups interact with her and thus with members from group A. This makes project a more attractive for members from group A, strengthening the incentives for players from group A to choose project a. This leads to more homophily. Conversely, if more players choose the group-preferred project, this strengthens the incentives of players with the group-preferred project to invest effort, as it increases their chances of meeting a player from their own group. This, in turn, further increases the chances for players with the group-preferred project of interacting with their own group, reinforcing the incentives to segregate. On the other hand, if effort is low, then the incentives to segregate are attenuated, as the probability of meeting similar others is small. This, in turn, reduces the incentives to invest effort.

As a result of this feedback loop, there are two regimes. If the cost c of effort is small relative to the benefits of coordination (as captured by v and Q_{in}), then players are willing to exert high effort, which in turn leads more players to choose the group-preferred project, further enhancing the incentives to invest effort. In that case, groups are segregated, and players are densely connected. Importantly, players with the group-preferred project face much stronger incentives to invest effort than players with the other project, as players with the group-preferred project have a high chance of interacting with heir own group.

On the other hand, if the effort cost c is high relative to the benefits of coordination, then the net benefit of interacting with others is small even if the society were fully segregated. In that case, choices are guided primarily by intrinsic preferences over projects, and the level of homophily is low. As a result, players face roughly the same incentives to invest effort, regardless of their project choice, and all players have approximately the same number of connections.

Hence, high levels of homophily go hand in hand with inequality in the number of connections that players have. The following result makes this precise:

Proposition 4.4. *Consider the limit where the number of players grows large (i.e., $|N| \to \infty$). In the unique introspective equilibrium, the distribution of connections of players with the group-preferred project first-order stochastically dominates the distribution of the number of connections of players with the other project. The difference in the expected number of connections of the players with the group-preferred project and the other project strictly increases with the level of homophily.*

This result follows directly from Proposition 4.2 and Theorem 3.13 of Bollobás, Janson, and Riordan (2007).[16]

There is substantial empirical evidence that, as our model predicts, more homogeneous groups have a higher level of social interactions (Alesina and La Ferrara, 2000) and that there is large variation in the number of connections that players have (Jackson, 2008). Furthermore, friendships are often biased towards own-group friendships, and larger groups form more friendships per capita (Currarini, Jackson, and Pin, 2009).

Our results put restrictions on the type of networks that can be observed. If the relative benefit v/c is limited and impulses are only weakly correlated within groups (i.e., Q_{in} close to $\frac{1}{2}$), networks are disconnected. Moreover, they feature low levels of homophily and limited variation in the number of connections. On the other hand, if the relative benefit v/c is high and impulses are strongly correlated within a group (i.e., Q_{in} close to 1), networks are dense. Moreover, they are characterized by high levels of homophily and a skewed distribution of the number of connections. Networks consist of a tightly connected core of players from one group, with a smaller periphery of players from the other group. This type of network is prevalent: many social and economic networks are dense, have a core-periphery structure with large variation in the number of connections, and feature high levels of homophily (Jackson, 2008).

5. Conclusions

Persistently high levels of homophily have long intrigued researchers. Rather than directly positing homophilous preferences, we derive them from a desire to reduce strategic uncertainty. Homophily emerges because players find it easier to predict the instinctive reactions of

[16] In fact, more can be said: the number of connections of a player with the group-preferred project converges to a Poisson random variable with parameter e^*, and the number of connections of players with the other project converges to a Poisson random variable with parameter $e^- < e^*$.

members of their own group. By providing microfoundations for homophilous preferences, we are able to derive novel testable hypotheses of how homophily varies with economic incentives. Our theory also puts restrictions on the types of networks that can be observed.

Our framework offers a versatile tool to study homophily in a broad range of environments. While we have restricted attention to a simple class of games to elucidate the main driving forces, the model can easily be adapted to encompass a richer class of games, so as to consider economically critical processes on networks such as information sharing.

We show that there may be homophily even if players have no direct preference for interacting with their own group. However, this does not, in itself, deliver an economic rationale for policies that reduce homophily. In our model, homophily is a by-product of socially valuable efforts to reduce strategic uncertainty, and homophily per se does not entail a welfare loss. A proper understanding of the root causes of homophily is thus also critical for welfare analysis. We leave this for future work.

Appendix A Equilibrium analysis

We compare the outcomes predicted using the introspective process to equilibrium predictions. As we show, the introspective process selects a correlated equilibrium of the game that has the highest level of homophily among the set of equilibria in which players' action depends on their signal, and thus maximizes the payoffs within this set.

We study the correlated equilibria of the extended game: in the first stage, players choose a project and are matched with players with the same project; and in the second stage, players play the coordination game with their partner. It is not hard to see that every introspective equilibrium is a correlated equilibrium. The game has more equilibria, though, even if we fix the signal structure. For example, in the coordination stage, the strategy profile under which all players choose the same fixed action regardless of their signal is a correlated equilibrium, as is the strategy profile under which half of the players in each group choose s^1 and the other half of the players choose s^2, or where players go against the action prescribed by their signal (i.e., choose s^2 if and only the signal is s^1). Given this, there is a plethora of equilibria for the extended game.

We restrict attention to equilibria in anonymous strategies, so that each player's equilibrium strategy depends only on his group, the project of the opponent he is matched with, and the signal he receives in the coordination game. In the coordination stage, we focus on equilibria in which players follow their signal. If all players follow their signal, following one's signal is a best response: for any probability p of interacting with a player of the own group, and any value w_j of a player's project, choosing action s^i having received signal i is a best response if and only if

$$\Big[pQ + (1-p)\cdot\tfrac{1}{2}\Big]\cdot v + w_j \geq \Big[p\cdot(1-Q_{in}) + (1-p)\cdot\tfrac{1}{2}\Big]\cdot v + w_j.$$

This inequality is always satisfied, as $Q_{in} > \frac{1}{2}$.

So, it remains to consider the matching stage. Suppose that $m^{A,a}$ and $m^{B,b}$ are the shares of A-players and B-players that choose projects a and b, respectively. Then, the probability that a player with project a belongs to group A is

$$p^{A,a} = \frac{m^{A,a}}{m^{A,a} + 1 - m^{B,b}};$$

similarly, the probability that a player with project b belongs to group B equals

$$p^{B,b} = \frac{m^{B,b}}{m^{B,b} + 1 - m^{A,a}}.$$

An A-player with intrinsic values $w_j^{A,a}$ and $w_j^{A,b}$ for the projects chooses project a if and only if

$$\Big[p^{A,a}Q_{in} + (1-p^{A,a})\cdot\tfrac{1}{2}\Big]\cdot v + w_j^{A,a} \geq \Big[(1-p^{B,b})\cdot Q_{in} + p^{B,b}\tfrac{1}{2}\Big]\cdot v + w_j^{A,b};$$

or, equivalently,

$$w_j^{A,a} - w_j^{A,b} \geq -(p^{A,a} + p^{B,b} - 1)\cdot\beta,$$

where we have defined $\beta := v\cdot(Q_{in} - \frac{1}{2})$. Similarly, a B-player with intrinsic values $w_j^{B,b}$ and $w_j^{B,a}$ chooses b if and only if

$$w_j^{B,b} - w_j^{B,a} \geq -(p^{A,a} + p^{B,b} - 1)\cdot\beta$$

In equilibrium, we must have that

$$\begin{aligned}\mathbb{P}\big(w_j^{A,a} - w_j^{A,b} \geq -(p^{A,a} + p^{B,b} - 1)\cdot\beta\big) &= m^{A,a}; \text{ and}\\ \mathbb{P}\big(w_j^{B,b} - w_j^{B,a} \geq -(p^{A,a} + p^{B,b} - 1)\cdot\beta\big) &= m^{B,b}.\end{aligned}$$

Because the random variables $w_j^{A,a} - w_j^{A,b}$ and $w_j^{B,b} - w_j^{B,a}$ have the same distribution (cf. Appendix C.2), it follows that $m^{A,a} = m^{B,b}$ and $p^{A,a} = p^{B,b}$ in equilibrium. Defining $p := p^{A,a}$ (and recalling the notation $\Delta_j := w_j^{A,a} - w_j^{A,b}$ from Appendix C.2), the equilibrium condition reduces to

$$\mathbb{P}(\Delta_j \geq -(2p-1)\cdot\beta) = p. \tag{A.1}$$

Thus, equilibrium strategies are characterized by a fixed point p of Equation (A.1).

It is easy to see that the introspective equilibrium characterized in Proposition 3.2 is an equilibrium. However, the game has more equilibria. The point $p = 0$ is a fixed point of (A.1) if and only if $\beta \geq 1$. In an equilibrium with $p = 0$, all A-players adopt project b, even if they have a strong intrinsic preference for project a, and analogously for B-players. In this case, the incentives for interacting with the own group, measured by β, are so large that they dominate any intrinsic preference.

But even if β falls below 1, we can have equilibria in which a minority of the players chooses the group-preferred project, provided that intrinsic preferences are not too strong. Specifically, it can be verified that there are equilibria with $p < \frac{1}{2}$ if and only if $\varepsilon \leq \frac{1}{2} - 2\beta(1-\beta)$. This condition is satisfied whenever ε is sufficiently small.

So, in general, there are multiple equilibria, and some equilibria in which players condition their action on their signal are inefficient as only a minority gets to choose the project they (intrinsically) prefer. Choosing a project is a coordination game, and it is possible to get stuck in an inefficient equilibrium. The introspective process described in Section 3 selects the payoff-maximizing equilibrium, with the largest possible share of players coordinating on the group-preferred project.

Importantly, the multiplicity of equilibria in the standard setting makes it difficult to derive unambiguous comparative statics. This is because as parameters are adjusted, the set of equilibria changes. Consider, for example, the effect of increasing the within-group similarity. As any introspective equilibrium is a correlated equilibrium, there is a correlated equilibrium where greater within-group similarity leads to more homophily (Corollary 3.3). But, varying the within-group similarity also changes the set of correlated equilibria. It is not hard to construct examples where increasing the within-group similarity gives rise to new (anonymous) correlated equilibria with lower levels of homophily.

Appendix B Signaling identity

Thus far, we have assumed that players can sort by choosing projects to sort. An alternative way in which individuals can bias the meeting process is by signaling their identity to others. Here, we assume that players can use markers, that is, observable attributes such as tattoos, to signal their identity.

There are two markers, a and b. Players first choose a marker, and are then matched to play the coordination game as described below. As before, each A-player has values $w_j^{A,a}$ and $w_j^{A,b}$ for markers a and b, drawn uniformly at random from $[0,1]$ and $[0, 1-2\varepsilon]$, respectively; and mutatis mutandis for a B-player. Thus, a is the *group-preferred marker* for group A, and b is the group-preferred marker for group B.

Players can now choose whether they want to interact with a player with an a- or a b-marker. Each player is chosen to be a *proposer* or a *responder* with equal probability, independently across players. Proposers can propose to play the coordination game to a responder. He chooses whether to propose to a player with an a- or a b-marker. If he chooses to propose with a player with an a-marker, he is matched uniformly at random with a responder with marker, and likewise if he chooses to propose to a player with a b-marker. A responder decides whether to accept or reject a proposal from a proposer, conditional on his own marker and the marker of the proposer.[17] Each player is matched exactly once.[18] Players' decision to propose or to accept/reject a proposal may depend on marker choice, but does not depend on players' identities or group membership, which is unobservable. If player j proposed to player j', and j' accepted j's proposal, then they play the coordination game in Section 2; if j's proposal was rejected by j', both get a payoff of zero.

Players again use introspection to decide on their action. At level 0, players choose the marker that they intrinsically prefer. Moreover, players propose to/accept proposals from anyone. At level 1, an A-player therefore has no incentive to choose a marker other than his intrinsically preferred marker, and thus chooses that marker. However, since at level 0, a slight majority of players with marker a belongs to group A, proposers from group A have an incentive to propose only to players with marker a, unless they have a strong intrinsic preference for marker b. Because players are matched only once, and because payoffs in the coordination game are nonnegative, a responder always accepts any proposal. The same holds, mutatis mutandis, for B-players.

We can prove an analogue of Proposition 3.2 for this setting:

Proposition B.1. [Equilibrium Characterization Marker Choice] *There is a unique introspective equilibrium of the extended game. In the unique equilibrium, players follow their impulse in the coordination game, and players' marker choices give rise to complete segregation ($h = \frac{1}{2}$) if and only if*

$$\beta \geq \tfrac{1}{2} - \varepsilon;$$

If segregation is not complete ($h < \frac{1}{2}$), then the level of homophily is given by:

$$\tfrac{1}{2} - \frac{1}{2 - 4\varepsilon}\left(1 - 2\varepsilon - \tfrac{1}{2} \cdot \beta\right)^2.$$

[17] So, a proposer only proposes to play, and a responder can only accept or reject a proposal. In particular, he cannot propose transfers. The random matching procedure assumed in Section 2 can be viewed as the reduced form of this process.

[18] Such a matching is particularly straightforward to construct when there are finitely many players. Otherwise, we can use the matching process of Alós-Ferrer (1999). The results continue to hold when players are matched a fixed finite number of times, or when there is discounting and players are sufficiently impatient. Without such restrictions, players have no incentives to accept a proposal from a player with the non-group preferred marker, leaving a significant fraction of the players unmatched.

In all cases, the fraction of players choosing the group-preferred marker exceeds the initial level (i.e., $h > \varepsilon$).

The proof is in the appendix. The equilibrium takes a similar form as when players can sort by choosing projects. The equilibrium level of homophily is always higher than the level of homophily based on preferences over markers. If the marginal benefit of interacting with the own group is sufficiently high, then there is full segregation.

Also the comparative statics are similar:

Corollary B.2. *The level of homophily h increases with the within-group similarity Q_{in} and with the coordination payoff v. Within-group similarity and economic incentives are complements: the level of homophily is high whenever the coordination payoff is high and the within-group similarity is close to 1.*

So, even if players cannot influence the probability of meeting similar others by locating in a particular neighborhood or joining a club, they can nevertheless associate preferentially with other members of their own group, provided that they can signal their identity. These results help explain why groups are often marked by seemingly arbitrary traits (Barth, 1969). Unlike in classical models of costly signaling, adopting a certain marker is *not* inherently more costly for one group than for another. The difference in signaling value of the markers across groups is endogenous in our model.

Appendix C Auxiliary results

C.1 Impulses

We characterize the probability that two players have the same impulse. Recall that, conditional on $\theta_A = 1$, an A-player has an impulse to play action s^1 with probability $q \in (\frac{1}{2}, 1)$. Likewise, conditional on $\theta_A = 2$, an A-player has an impulse to play action s^2 with probability q. Analogous statements apply to B-players. The states θ_A and θ_B are independent. The following result from Kets and Sandroni (2015) characterizes the probability that two players have the same impulse.

Lemma C.1. [**Kets and Sandroni (2015)**] *Let $q \in (\frac{1}{2}, 1)$ be the probability that a player of group A has the impulse to choose s^1 conditional on $\theta_A = 1$, and analogously for group B. Then:*

(a) the probability that two distinct A-players have the same impulse is $Q_{in} := q^2 + (1-q)^2 \in (\frac{1}{2}, 1)$;

(b) the probability that two distinct A-players have the impulse to play s^1 is equal to $\frac{1}{2}Q_{in}$;

(c) the conditional probability that an A-player j has the impulse to play action s^1 given that another A-player j' has the impulse to play action s^1 is equal to Q_{in};

(d) the probability that an A-player and a B-player have the same impulse is $Q_{out} = \frac{1}{2}$.

(e) the probability that an A-player and a B-player have the impulse to play s^1 is equal to $\frac{1}{2}Q_{out}$;

(f) the conditional probability that an A-player j has the impulse to play action s^1 given that a B-player j' has the impulse to play s^1 is equal to Q_{out};

C.2 Intrinsic preferences

We denote the values of an A-player j for projects a and b are denoted by $w_j^{A,a}$ and $w_j^{A,b}$, respectively; likewise, the values of a B-player for projects b and a are $w_j^{B,b}$ and $w_j^{B,a}$, respectively. As noted in the main text, the values $w_j^{A,a}$ and $w_j^{A,b}$ are drawn from the uniform distribution on $[0, 1]$ and $[0, 1 - 2\varepsilon]$, respectively. Likewise, $w_j^{B,b}$ and $w_j^{B,a}$ are uniformly distributed on $[0, 1]$ and $[0, 1-2\varepsilon]$. All values are drawn independently (across players, projects, and groups). So, players in group A (on average) intrinsically prefer project a (in the sense of first-order stochastic dominance) over project b; see Figure 4. Likewise, on average, players in group B have an intrinsic preference for b.

Given that the values are uniformly and independently distributed, the distribution of the difference $w_j^{A,a} - w_j^{B,a}$ in values for an A-player is given by the so-called trapezoidal distribution. That is, if we define $x := 1-2\varepsilon$, we can define the tail distribution $H_\varepsilon(y) := \mathbb{P}(w_j^{A,a} - w_j^{A,b} \geq y)$ by

$$H_\varepsilon(y) = \begin{cases} 1 & \text{if } y < -(1-2\varepsilon); \\ 1 - \frac{1}{2-4\varepsilon} \cdot (1 - 2\varepsilon + y)^2 & \text{if } y \in [-(1-2\varepsilon), 0); \\ 1 - \frac{1}{2} \cdot (1 - 2\varepsilon) - y & \text{if } y \in [0, 2\varepsilon); \\ \frac{1}{4(\frac{1}{2}-\varepsilon)} \cdot (1-y)^2 & \text{if } y \in [2\varepsilon, 1]; \\ 0 & \text{otherwise.} \end{cases}$$

By symmetry, the probability $\mathbb{P}(w_j^{B,b} - w_j^{B,a} \geq y)$ that the difference in values for the B-player is at least y is also given by $H_\varepsilon(y)$. So, we can identify $w_j^{A,a} - w_j^{A,b}$ and $w_j^{B,b} - w_j^{B,a}$ with the same random variable, denoted Δ_j, with tail distribution $H_\varepsilon(\cdot)$; see Figure 5.

The probability that A-players prefer the a-project, or, equivalently, the share of A-players that intrinsically prefer a (i.e., $w_j^{A,a} - w_j^{A,b} > 0$), is $1 - \frac{1}{2}x = \frac{1}{2} + \varepsilon$, and similarly for the B-players and project b.

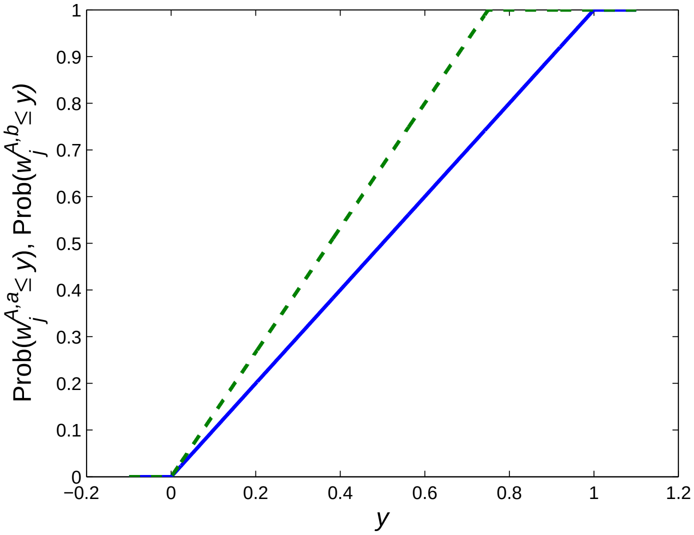

Figure 4: The cumulative distribution functions of $w_j^{A,a}$ (solid line) and $w_j^{A,b}$ (dashed line) for $x = 0.75$.

Appendix D Proofs

D.1 Proof of Proposition 2.2

The proof follows from Lemma 3.2 of Kets and Sandroni (2015). We give a separate proof here to give insight in the driving forces. By assumption, a player chooses action s^i at level 0 if and only if his initial impulse is $i = 1, 2$. For $k > 0$, assume, inductively, that at level $k - 1$, a player chooses s^i if and only if his initial impulse is i. Consider level k, and suppose a player's impulse is i. Choosing s^i is the unique best response for him if the expected payoff from choosing s^i is strictly greater than the expected payoff from choosing the other action $s^j \neq s^i$. That is, if we write $j \neq i$ for the alternate impulse, s^i is the unique best response for the player if

$$p \cdot v \cdot \mathbb{P}(i \mid i) + (1-p) \cdot v \cdot \mathbb{P}(i) > p \cdot v \cdot \mathbb{P}(j \mid i) + (1-p) \cdot v \cdot \mathbb{P}(j) \cdot v,$$

where $\mathbb{P}(m \mid i)$ is the conditional probability that the impulse of a player from the same group is $m = 1, 2$ given that the player's own impulse is i, and $\mathbb{P}(m)$ is the probability that a player from the other group has received signal m. Using that $\mathbb{P}(m) = \frac{1}{2}$, $\mathbb{P}(i \mid i) = q^2 + (1-q)^2$ and $\mathbb{P}(j \mid i) = 1 - q^2 - (1-q)^2$, and rearranging, we find that this holds if and only if

$$p(q^2 + (1-q)^2) > p(1 - q^2 - (1-q)^2),$$

and this holds for every $p > 0$, since $q^2 + (1-q)^2 > \frac{1}{2}$. This shows that at each level, it is optimal for a player to follow his impulse. So, in the unique introspective equilibrium, every

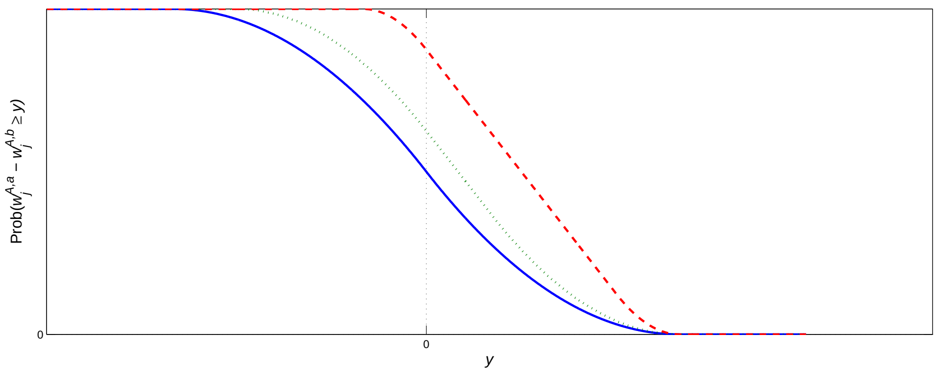

Figure 5: The probability that $w_j^{A,a} - w_j^{A,b}$ is at least y, as a function of y, for $\varepsilon = 0$ (solid line); $\varepsilon = 0.125$ (dotted line); and $\varepsilon = 0.375$ (dashed line).

player follows his impulse. □

D.2 Proof of Lemma 3.1

At level 0, players choose the project that they intrinsically prefer. So, the share of players that choose project a that belong to group A is

$$p_0^a = \frac{\frac{1}{2}+\varepsilon}{\frac{1}{2}+\varepsilon+(1-(\frac{1}{2}+\varepsilon))} = \tfrac{1}{2} + \varepsilon.$$

Likewise, the share of players that choose project b that belong to group B is $p_0^b = \frac{1}{2}+\varepsilon$. Also, recall that $x := 1 - 2\varepsilon$ (Appendix C.2).

Recall that marginal benefit of interacting with the own group is

$$\beta := v \cdot (Q_{in} - Q_{out}).$$

As $Q_{in} > Q_{out}$, the marginal benefit of interacting with the own group is positive. We show that the sequence $\{p_k^\pi\}_k$ is (weakly) increasing and bounded for every project π.

At higher levels, players choose projects based on their intrinsic values for the project as well as the coordination payoff they expect to receive at each project. Suppose that a share p_{k-1}^a of players with project a belong to group A, and likewise for project b and group B. Then, the probability that an A-player with project a is matched with a player of the own group is p_{k-1}^a, and the probability that a B-player with project a is matched with a player of the own group is $1 - p_{k-1}$. Applying Proposition 2.2 (with $\hat{p} = p_{k-1}$ and $\hat{p} = 1 - p_{k-1}$) shows that both A-players and B-players with project a follow their signal in the coordination game, and similarly for the A- and B-players with project b.

So, for every $k > 0$, given p^a_{k-1}, a player from group A chooses project a if and only if

$$\Big[p^a_{k-1} \cdot Q_{in} + (1 - p^a_{k-1}) \cdot Q_{out}\Big] \cdot v + w^{A,a}_j \geq \Big[(1 - p^a_{k-1}) \cdot Q_{in} + p^a_{k-1} \cdot Q_{out}\Big] \cdot v + w^{A,b}_j.$$

This inequality can be rewritten as

$$w^{A,a}_j - w^{A,b}_j \geq -(2p^a_{k-1} - 1) \cdot \beta, \tag{D.1}$$

and the share of A-players for whom this holds is

$$p^a_k := H_\varepsilon\big(-(2p_{k-1} - 1) \cdot \beta\big),$$

where we have used the expression for the tail distribution H_ε from Appendix C.2. The same law of motion holds, of course, if a is replaced with b and A is replaced with B.

Fix a project π. Notice that $-(2p^\pi_0 - 1) \cdot \beta < 0$. We claim that $p^\pi_1 \geq p^\pi_0$ and that $p^\pi_1 \in (\frac{1}{2}, 1]$. By the argument above,

$$\begin{aligned} p^\pi_1 &= \mathbb{P}(w^{A,a}_j - w^{A,b}_j \geq -(2p^\pi_0 - 1) \cdot \beta) \\ &= H_\varepsilon(-(2p^\pi_0 - 1) \cdot \beta) \\ &= \begin{cases} 1 - \frac{1}{2-4\varepsilon} \cdot (1 - 2\varepsilon - (2p^\pi_0 - 1) \cdot \beta)^2 & \text{if } (2p^\pi_0 - 1) \cdot \beta \leq 1 - 2\varepsilon; \\ 1 & \text{if } (2p^\pi_0 - 1) \cdot \beta > 1 - 2\varepsilon; \end{cases} \end{aligned}$$

where we have used the expression for the tail distribution $H_\varepsilon(y)$ from Appendix C.2. If $(2p^\pi_0 - 1) \cdot \beta > 1 - 2\varepsilon$, the result is immediate, so suppose that $(2p^\pi_0 - 1) \cdot \beta \leq 1 - 2\varepsilon$. We need to show that

$$1 - \tfrac{1}{2-4\varepsilon} \cdot (1 - 2\varepsilon - (2p^\pi_0 - 1) \cdot \beta)^2 \geq p^\pi_0.$$

Rearranging and using that $p^\pi_0 \in (\frac{1}{2}, 1]$, we see that this holds if and only if

$$(2p^\pi_0 - 1) \cdot \beta \leq 2 \cdot (1 - 2\varepsilon).$$

But this holds because $(2p^\pi_0 - 1) \cdot \beta \leq 1 - 2\varepsilon$ and $1 - 2\varepsilon \geq 0$. Note that the inequality is strict whenever $\beta < 1 - 2\varepsilon$, so that $p^\pi_1 > p^\pi_0$ in that case.

For $k > 1$, suppose, inductively, that $p^\pi_{k-1} \geq p^\pi_{k-2}$ and that $p^\pi_{k-1} \in (\frac{1}{2}, 1]$. By a similar argument as above,

$$p^\pi_k = \begin{cases} 1 - \frac{1}{2-4\varepsilon} \cdot (1 - 2\varepsilon - (2p^\pi_{k-1} - 1) \cdot \beta)^2 & \text{if } (2p^\pi_{k-1} - 1) \cdot \beta \leq 1 - 2\varepsilon; \\ 1 & \text{if } (2p^\pi_{k-1} - 1) \cdot \beta > 1 - 2\varepsilon. \end{cases}$$

Again, if $(2p^\pi_{k-1} - 1) \cdot \beta > 1 - 2\varepsilon$, the result is immediate, so suppose $(2p^\pi_{k-1} - 1) \cdot \beta \leq 1 - 2\varepsilon$. We need to show that

$$1 - \tfrac{1}{2-4\varepsilon} \cdot (1 - 2\varepsilon - (2p^\pi_{k-1} - 1) \cdot \beta)^2 \geq p^\pi_{k-1},$$

or, equivalently,

$$2 \cdot (1-2\varepsilon) \cdot (1-p^{\pi}_{k-1}) \geq (1-2\varepsilon-(2p^{\pi}_{k-1}-1) \cdot \beta)^2.$$

By the induction hypothesis, $p^{\pi}_{k-1} \geq p^{\pi}_0$, so that $1-2\varepsilon \geq 2-2p^{\pi}_{k-1}$. Using this, we have that $2 \cdot (1-2\varepsilon) \cdot (1-p^{\pi}_{k-1}) \geq 4 \cdot (1-p^{\pi}_{k-1})^2$. Moreover,

$$(1-2\varepsilon-(2p^{\pi}_{k-1}-1) \cdot \beta)^2 \leq 4 \cdot (1-p^{\pi}_{k-1})^2 - 2\beta(1-2\varepsilon)(2p^{\pi}_{k-1}-1) + (2p^{\pi}_{k-1}-1)^2\beta^2.$$

So, it suffices to show that

$$4 \cdot (1-p^{\pi}_{k-1})^2 \geq 4 \cdot (1-p^{\pi}_{k-1})^2 - 2\beta(1-2\varepsilon)(2p^{\pi}_{k-1}-1) + (2p^{\pi}_{k-1}-1)^2\beta^2.$$

The above inequality holds if and only if

$$(2p^{\pi}_{k-1}-1)\beta \leq 2 \cdot (1-2\varepsilon),$$

and this is true since $(2p^{\pi}_{k-1}-1) \cdot \beta \leq 1-2\varepsilon$.

So, the sequence $\{p^{\pi}_k\}_k$ is weakly increasing and bounded when $\beta > 0$. It now follows from the monotone sequence convergence theorem that the limit p^{π} exists. The argument clearly does not depend on the project π, so we have $p^a = p^b$. □

D.3 Proof of Proposition 3.2

Recall that the marginal benefit of interacting with the own group is $\beta > 0$. The first step is to characterize the limiting fraction p, and show that $p > \frac{1}{2} + \varepsilon$. By the proof of Lemma 3.1, we have $p_k \geq p_{k-1}$ for all k. By the monotone sequence convergence theorem, $p = \sup_k p_k$, and by the inductive argument, $p \in (\frac{1}{2}+\varepsilon, 1]$. It is easy to see that $p = 1$ if and only if $H_\varepsilon(-(2 \cdot 1 - 1) \cdot \beta) = 1$, which holds if and only if $\beta \geq 1-2\varepsilon$.

So suppose that $\beta < 1-2\varepsilon$, so that $p < 1$. Again, $p = H_\varepsilon(-(2p-1) \cdot \beta)$, or, using the expression from Appendix C.2,

$$p = 1 - \tfrac{1}{2-4\varepsilon} \cdot (1-2\varepsilon-(2p-1) \cdot \beta)^2.$$

It will be convenient to substitute $x = 1-2\varepsilon$ for ε, so that we are looking for the solution of

$$p = 1 - \tfrac{1}{2x} \cdot (x-(2p-1) \cdot \beta)^2. \tag{D.2}$$

Equation (D.2) has two roots,

$$r_1 = \tfrac{1}{2} + \tfrac{1}{4\beta^2}\Big((2\beta-1) \cdot x + \sqrt{4\beta^2 x - (4\beta-1) \cdot x^2}\Big)$$

and

$$r_2 = \tfrac{1}{2} + \tfrac{1}{4\beta^2}\Big((2\beta - 1) \cdot x - \sqrt{4\beta^2 x - (4\beta - 1) \cdot x^2}\Big).$$

We first show that r_1 and r_2 are real numbers, that is, that $4\beta^2 x - (4\beta - 1) \cdot x^2 \geq 0$. Since $x > 0$, this is the case if and only if $4\beta \geq (4\beta - 1) \cdot x$. This holds if $\beta \leq \frac{1}{4}$, so suppose that $\beta > \frac{1}{4}$. We need to show that

$$x \leq \frac{4\beta^2}{4\beta - 1}.$$

Since the right-hand side achieves its minimum at $\beta = \frac{1}{2}$, it suffices to show that $x \leq (4 \cdot (\frac{1}{2})^2)/(4 \cdot \frac{1}{2} - 1) = 1$. But this holds by definition. It follows that r_1 and r_2 are real numbers.

We next show that $r_1 > \frac{1}{2}$, and $r_2 < \frac{1}{2}$. This implies that $p = r_1$, as $p = \sup_k p_k > p_0 > \frac{1}{2}$.

It suffices to show that $4\beta^2 x - (4\beta - 1) \cdot x^2 > (1 - 2\beta)^2 x^2$. This holds if and only if $\beta > (2 - \beta) \cdot x$. Recalling that $\beta \leq 1 - 2\varepsilon < 1$ by assumption, we see that this inequality is satisfied. We conclude that $p = r_1$ when $\beta > 0$. □

D.4 Proof of Corollary 3.3

It is straightforward to verify that the derivative of p with respect to β is positive whenever $p < 1$ (and 0 otherwise). It then follows from the chain rule that the derivatives of p with respect to v and Q_{in} are both positive for any $p < 1$ (and 0 otherwise). □

D.5 Proof of Lemma 4.1

Recall that at level 0, players invest effort $e_0 > 0$ in socializing. Moreover, they choose project a if and only if they intrinsically prefer project a over project b. It follows from the distribution of the intrinsic values (Appendix C.2) that the number $N_0^{A,a}$ of A-players with project a at level 0 follows the same distribution as the number $N_0^{B,b}$ of B-players with project b at level 0; similarly, the number $N_0^{A,a}$ of A-players with project b at level 0 has the same distribution as the number $N_0^{B,a}$ of B-players with project a at level 0. Let N_0^D and N_0^M be random variables with the same distribution as $N_0^{A,a}$ and $N_0^{B,a}$, respectively (where D stands for "dominant group" and M stands for "minority group"; the motivation for this terminology is that a slight majority of the players with an intrinsic preference for project a belongs to group A).

Conditional on N_0^D and N_0^M, the expected utility of project a to an A-player at level 1 is[19]

$$v \cdot \left[\frac{e_j \cdot N_0^D \cdot e_0 \cdot Q_{in} + e_j \cdot N_0^M \cdot e_0 \cdot \frac{1}{2}}{N_0^D \cdot e_0 + N_0^M \cdot e_0}\right] + w_j^{A,a} - \frac{ce_j}{2}$$

[19] If $N_0^D = N_0^M = 0$, then the expected benefit from networking is 0. In that case, the player's expected utility is thus $w_j^{A,a} - \frac{ce_j}{2}$. A similar statement applies at higher levels.

if he invests effort e_j and his intrinsic value for project a is $w_j^{A,a}$. Likewise, conditional on N_0^D and N_0^M, the expected utility of project b to an A-player at level 1 is

$$v \cdot \left[\frac{e_j \cdot N_0^M \cdot e_0 \cdot Q_{in} + e_j \cdot N_0^D \cdot e_0 \cdot \frac{1}{2}}{N_0^D \cdot e_0 + N_0^M \cdot e_0}\right] + w_j^{A,b} - \frac{ce_j}{2}$$

if he invests effort e_j and his intrinsic value for project b is $w_j^{A,b}$. Taking expectations over N_0^D and N_0^M, it follows from the first-order conditions that the optimal effort levels for an A-player at level 1 with projects a and b are given by

$$\begin{aligned} e_1^{A,a} &= \left(\frac{v}{c}\right) \cdot \mathbb{E}\left[\frac{N_0^D \cdot e_0 \cdot Q_{in} + N_0^M \cdot e_0 \cdot \frac{1}{2}}{N_0^D \cdot e_0 + N_0^M \cdot e_0}\right]; \text{ and} \\ e_1^{A,b} &= \left(\frac{v}{c}\right) \cdot \mathbb{E}\left[\frac{N_0^M \cdot e_0 \cdot Q_{in} + N_0^D \cdot e_0 \cdot \frac{1}{2}}{N_0^D \cdot e_0 + N_0^M \cdot e_0}\right]; \end{aligned}$$

respectively, independent of the intrinsic values. It can be checked that the optimal effort levels $e_1^{B,a}$ and $e_1^{B,b}$ for a B-player at level 1 with projects a and b are equal to $e_1^{A,b}$ and $e_1^{A,a}$, respectively. It will be convenient to define $e_1^D := e_1^{A,a} = e_1^{B,b}$ and $e_1^M := e_1^{A,b} = e_1^{B,a}$. We claim that $e_1^D > e_1^M$. To see this, note that N_0^D is binomially distributed with parameters $|N|$ and $p_0 := \frac{1}{2}+\varepsilon > \frac{1}{2}$ (the probability that a player has an intrinsic preference for the group-preferred project) and that N_0^M is binomially distributed with parameters $|N|$ and $1 - p_0 < \frac{1}{2}$. If we define

$$\begin{aligned} g_1^D(N_0^D, N_0^M, e_0) &:= \left(\frac{v}{c}\right) \cdot \left(\frac{N_0^D \cdot e_0 \cdot Q_{in} + N_0^M \cdot e_0 \cdot \frac{1}{2}}{N_0^D \cdot e_0 + N_0^M \cdot e_0}\right); \text{ and} \\ g_1^M(N_0^D, N_0^M, e_0) &:= \left(\frac{v}{c}\right) \cdot \left(\frac{N_0^M \cdot e_0 \cdot Q_{in} + N_0^D \cdot e_0 \cdot \frac{1}{2}}{N_0^D \cdot e_0 + N_0^M \cdot e_0}\right); \end{aligned}$$

so that e_1^D and e_1^M are just the expectations of g_1^D and g_1^D, respectively, then the result follows immediately from the fact that N_0^D first-order stochastically dominates N_0^M, as g_1^D is (strictly) increasing in N_0^D and (strictly) decreasing in N_0^M, and g_1^M is decreasing in N_0^D and increasing in N_0^M (again, strictly).

Substituting the optimal effort levels e_1^D and e_1^M into the expression for the expected utility for each project shows that the maximal expected utility of an A-player at level 1 of projects a and b is given by

$$\begin{aligned} &\frac{c}{2}(e_1^D)^2 + w_j^{A,a}; \text{ and} \\ &\frac{c}{2}(e_1^M)^2 + w_j^{A,b}; \end{aligned}$$

respectively. At level 1, an A-player therefore chooses project a if and only if

$$w_j^{A,a} - w_j^{A,b} \geq -\frac{c}{2}\big((e_1^D)^2 - (e_1^M)^2\big).$$

The analogous argument shows that a B-player chooses project b at level 1 if and only if

$$w_j^{B,b} - w_j^{B,a} \geq -\frac{c}{2}\big((e_1^D)^2 - (e_1^M)^2\big).$$

Since $w_j^{A,a} - w_j^{A,b}$ and $w_j^{B,b} - w_j^{B,a}$ both have tail distribution $H_\varepsilon(\cdot)$ (Appendix C.2), the probability that an A-player chooses project a (or, that a B-player chooses project b) is

$$p_1 := H_\varepsilon\left(-\frac{c}{2}\big((e_1^D)^2 - (e_1^M)^2\big)\right).$$

Since $e_1^D > e_1^M$, we have $p_1 > p_0$. Note that both the number $N_1^{A,a}$ of A-players at level 1 with project a and the number $N_1^{B,b}$ of B-players at level 1 with project b are binomially distributed with parameters $|N|$ and $p_1 > \frac{1}{2}$; the number $N_1^{A,a}$ of A-players at level 1 with project b and the number $N_1^{B,a}$ of B-players at level 1 with project a are both binomially distributed with parameters $|N|$ and $1 - p_1$. Let N_1^D and N_1^M be random variables that are binomially distributed with parameters $(|N|, p_1)$ and $(|N|, 1 - p_1)$, respectively, so that the distribution of N_1^D first-order stochastically dominates the distribution of N_1^M.

Note that while $N_1^{A,a}$ and $N_1^{A,a}$ are clearly not independent (as $N_1^{A,a} + N_1^{A,a} = N$), $N_1^{A,a}$ and $N_1^{B,a}$ are independent (and similarly if we replace $N_1^{A,a}$, $N_1^{A,a}$, and $N_1^{B,a}$ with $N_1^{B,b}$, $N_1^{B,a}$, and $N_1^{A,a}$, respectively). When we take expectations over the number of players from different groups with a given project (e.g., $N_1^{A,a}$ and $N_1^{B,a}$) to calculate optimal effort levels, we therefore do not have to worry about correlations between the random variables. A similar comment applies to levels $k > 1$.

Finally, it will be useful to note that

$$e_1^D + e_1^M = \frac{v}{c}(Q_{in} + \tfrac{1}{2}).$$

Both e_1^D and e_1^M are positive, as they are proportional to the expectation of a nonnegative random variable (with a positive probability on positive realizations), and we have

$$e_1^D - e_1^M > e_0^D - e_0^M = 0,$$

where $e_0^D = e_0^M = e_0$ are the effort choices at level 0.

For $k > 1$, assume, inductively, that the following hold:

- we have $p_{k-1} \geq p_{k-2}$;
- the number $N_{k-1}^{A,a}$ of A-players with project a at level $k-1$ and the number $N_{k-1}^{B,b}$ of B-players with project b at level $k-1$ are binomially distributed with parameters $|N|$ and p_{k-1};

- the number $N^{A,a}_{k-1}$ of A-players with project b at level $k-1$ and the number $N^{B,a}_{k-1}$ of B-players with project a at level $k-1$ are binomially distributed with parameters $|N|$ and $1-p_{k-1}$;

- for every level $m \leq k-1$, the optimal effort level at level m for all A-players with project a and for all B-players with project b is equal to e^D_m;

- for every level $m \leq k-1$, the optimal effort level at level m for all A-players with project b and for all B-players with project a is equal to e^M_m;

- we have $e^D_{k-1} > e^M_{k-1} > 0$ for $k \geq 2$;

- we have $e^D_{k-1} - e^M_{k-1} \geq e^D_{k-2} - e^M_{k-2}$.

We write N^D_{k-1} and N^M_{k-1} for random variables that are binomially distributed with parameters $(|N|, p_{k-1})$ and $(|N|, 1-p_{k-1})$, respectively.

By a similar argument as before, it follows that the optimal effort level for an A-player that chooses project a or for a B-player that chooses b is

$$e^D_k := \left(\frac{v}{c}\right) \cdot \mathbb{E}\left[\frac{N^D_{k-1} \cdot e^D_{k-1} \cdot Q_{in} + N^M_{k-1} \cdot e^M_{k-1} \cdot \frac{1}{2}}{N^D_{k-1} \cdot e^D_{k-1} + N^M_{k-1} \cdot e^M_{k-1}}\right],$$

and that the optimal effort level for an A player that chooses project b or for a B-player that chooses a is

$$e^M_k := \left(\frac{v}{c}\right) \cdot \mathbb{E}\left[\frac{N^M_{k-1} \cdot e^M_{k-1} \cdot Q_{in} + N^D_{k-1} \cdot e^D_{k-1} \cdot \frac{1}{2}}{N^D_{k-1} \cdot e^D_{k-1} + N^M_{k-1} \cdot e^M_{k-1}}\right].$$

Again, it is easy to verify that

$$e^D_k + e^M_k = \frac{v}{c}(Q_{in} + \tfrac{1}{2}). \tag{D.3}$$

We claim that $e^D_k \geq e^D_{k-1}$ (so that $e^M_k \leq e^M_{k-1}$). It then follows from the induction hypothesis that $e^D_k > e^M_k$ and that $e^D_k - e^M_k \geq e^D_{k-1} - e^M_{k-1}$.

To show this, recall that for $m = 1, \ldots, k-1$, we have that $e^D_m > e^M_m$ and $e^D_m + e^M_m = \frac{v}{c}(Q_{in} + \frac{1}{2})$. Define

$$\begin{aligned} g^D_{k-1}(N^D_{k-2}, N^M_{k-2}, e^D_{k-2}) &:= \left(\frac{v}{c}\right) \cdot \left(\frac{N^D_{k-2} \cdot e^D_{k-2} \cdot Q_{in} + N^M_{k-2} \cdot e^M_{k-2} \cdot \frac{1}{2}}{N^D_{k-2} \cdot e^D_{k-2} + N^M_{k-2} \cdot e^M_{k-2}}\right) \\ g^D_k(N^D_{k-1}, N^M_{k-1}, e^D_{k-1}) &:= \left(\frac{v}{c}\right) \cdot \left(\frac{N^D_{k-1} \cdot e^D_{k-1} \cdot Q_{in} + N^M_{k-1} \cdot e^M_{k-1} \cdot \frac{1}{2}}{N^D_{k-1} \cdot e^D_{k-1} + N^M_{k-1} \cdot e^M_{k-1}}\right) \end{aligned}$$

so that e^D_{k-1} and e^D_k are just proportional to the expectation of g^D_{k-1} and g^D_k (over N^D_{k-1} and N^M_{k-1}), respectively, analogous to before. It is easy to verify that $g^D_k(N^D_{k-1}, N^M_{k-1}, e^D_{k-1}) \geq$

$g_k^D(N_{k-1}^D, N_{k-1}^M, e_{k-1}^M)$. Consequently,

$$\begin{aligned} e_k^D &\geq \left(\frac{v}{c}\right) \cdot \mathbb{E}\left[\frac{N_{k-1}^D \cdot e_{k-1}^M \cdot Q_{in} + N_{k-1}^M \cdot e_{k-1}^M \cdot \frac{1}{2}}{N_{k-1}^D \cdot e_{k-1}^M + N_{k-1}^M \cdot e_{k-1}^M}\right] \\ &= \left(\frac{v}{c}\right) \cdot \mathbb{E}\left[\frac{N_{k-1}^D \cdot Q_{in} + N_{k-1}^M \cdot \frac{1}{2}}{N_{k-1}^D + N_{k-1}^M}\right]. \end{aligned}$$

Using that g_k^D is decreasing in its second argument, and that the distribution of N_{k-2}^M first-order stochastically dominates the distribution of N_{k-1}^M, we have

$$e_k^D \geq \left(\frac{v}{c}\right) \cdot \mathbb{E}\left[\frac{N_{k-1}^D \cdot Q_{in} + N_{k-2}^M \cdot \frac{1}{2}}{N_{k-1}^D + N_{k-2}^M}\right]. \tag{D.4}$$

From the other direction, use that $g_{k-1}^D(N_{k-2}^D, N_{k-2}^M, e_{k-2}^M) \leq g_{k-1}^D(N_{k-2}^D, N_{k-2}^M, e_{k-1}^D)$ to obtain

$$e_{k-1}^D \leq \left(\frac{v}{c}\right) \cdot \mathbb{E}\left[\frac{N_{k-2}^D \cdot e_{k-2}^D \cdot Q_{in} + N_{k-2}^M \cdot e_{k-2}^D \cdot \frac{1}{2}}{N_{k-2}^D \cdot e_{k-2}^D + N_{k-2}^M \cdot e_{k-2}^D}\right].$$

Using that g_{k-1}^D is increasing in its first argument, and that the distribution of N_{k-1}^D first-order stochastically dominates the distribution of N_{k-2}^D, we obtain

$$e_{k-1}^D \leq \left(\frac{v}{c}\right) \cdot \mathbb{E}\left[\frac{N_{k-1}^D \cdot Q_{in} + N_{k-1}^M \cdot \frac{1}{2}}{N_{k-1}^D + N_{k-1}^M}\right]. \tag{D.5}$$

The result now follows by comparing Equations (D.4) and (D.5). Also, using that g_k^D is increasing and decreasing in its first and second argument, respectively, we have that

$$\begin{aligned} e_k^D &\geq \left(\frac{v}{c}\right) \cdot \mathbb{E}\left[\frac{N \cdot e_{k-1}^M \cdot \frac{1}{2}}{N \cdot e_{k-1}^M}\right] = \frac{v}{2c} \\ e_k^D &\leq \left(\frac{v}{c}\right) \cdot \mathbb{E}\left[\frac{N \cdot e_{k-1}^D \cdot Q_{in}}{N \cdot e_{k-1}^D}\right] = \frac{v \cdot Q_{in}}{c}, \end{aligned}$$

and it follows from (D.3) that $e_k^D, e_k^M \in [\frac{v}{2c}, \frac{v \cdot Q_{in}}{c}]$.

By a similar argument as before, the probability at level k that an A-player chooses project a (or, that a B-player chooses project b) is

$$p_k := H_\varepsilon\left(-\frac{c}{2}\big((e_k^D)^2 - (e_k^M)^2\big)\right).$$

Hence, the number $N_k^{A,a}$ of A-players with project a (or, the number $N^{B,b}$ of B-players with project b) at level k is a binomially distributed random variable N_k^D with parameters $|N|$ and p_k. Similarly, the number $N_k^{A,a}$ of A-players with project b (or, the number $N^{B,a}$ of B-players with project a) at level k is a binomially distributed random variable with parameters $|N|$ and $1 - p_k$.

Using that $e_k^D - e_k^M \geq e_{k-1}^D - e_{k-1}^M > 0$, and Equation (D.3) again, it follows that $(e_k^D)^2 - (e_k^M)^2 \geq (e_{k-1}^D)^2 - (e_{k-1}^M)^2 > 0$, it follows that $p_k \geq p_{k-1}$, and the induction is complete.

We thus have that the sequences p_0, p_1, p_2 and $e_1^D, e_2^D, \ldots$ are monotone and bounded, so that by the monotone convergence theorem, their respective limits $p := \lim_{k\to\infty} p_k$ and $e^D := \lim_{k\to\infty} e_k^D$ exist (as does $e^M := \lim_{k\to\infty} e_k^M = \frac{v}{c}(Q_{in} + \frac{1}{2}) - e^D$). □

D.6 Proof of Proposition 4.2

Recall the definitions from the proof of Lemma 4.1. It is straightforward to check that the random variables N_k^D and N_k^M converge in distribution to a binomially distributed random variable N^D with parameters $|N|$ and p and a binomially distributed random variable N^M with parameters $|N|$ and $1 - p$. It then follows from continuity and the Helly-Bray theorem that e^D satisfies

$$e^D = \left(\frac{v}{c}\right) \cdot \mathbb{E}\left[\frac{N^D \cdot e^D \cdot Q_{in} + N^M \cdot e^M \cdot \frac{1}{2}}{N^D \cdot e^D + N_{k-2}^M \cdot e^M}\right].$$

where the expectation is taken over N^D and N^M, so that e^D is a function of p. Also, by continuity, the limit p satisfies

$$p = H_\varepsilon\left(-\frac{c}{2}\left((e^D)^2 - (e^M)^2\right)\right).$$

By the proof of Lemma 4.1, we have $0 < e^M < e^D < \frac{v}{c}(Q_{in} + \frac{1}{2})$. Moreover, $e^D + e^M = \frac{v}{c}(Q_{in} + \frac{1}{2})$.

It remains to show that the equilibrium is unique (after all, the equations above could have multiple solutions). Define

$$h^D(e^D) := \left(\frac{v}{c}\right) \cdot \mathbb{E}\left[\frac{N^D \cdot e^D \cdot Q_{in} + N^M \cdot e^M \cdot \frac{1}{2}}{N^D \cdot e^D + N^M \cdot e^M}\right],$$

so that $e^D = h^D(e^D)$ in the introspective equilibrium.[20] Since $e^D + e^M = \frac{v}{c}(Q_{in} + \frac{1}{2})$ and $e^M > 0$, we have $e^D \in (0, \frac{v}{c}(Q_{in} + \frac{1}{2}))$. It is easy to check that $\lim_{e^D \downarrow 0} h^D(e^D) = \frac{v}{2c} > 0$ and that $\lim_{e^D \uparrow \frac{v}{c}(Q_{in}+\frac{1}{2})} h^D(e^D) = \frac{vQ}{c} < \frac{v}{c}(Q_{in} + \frac{1}{2})$. So, to show that there is a unique introspective equilibrium, it suffices to show that $h^D(e^D)$ is increasing and concave.

To show that $h^D(e^D)$ is increasing, define

$$g^D(N^D, N^M, e^D) := \frac{N^D \cdot e^D \cdot Q_{in} + N^M \cdot e^M \cdot \frac{1}{2}}{N^D \cdot e^D + N_{k-2}^M \cdot e^M},$$

[20] As before, the expectation is taken over N^D, N^M such that $N^D > 0$ or $N^M > 0$.

so that $h^D(e^D)$ is proportional to the expectation of g^D over N^D and N^M, as before. It is easy to verify that $g^D(N^D, N^M, e^D)$ is increasing in e^D for all N^D and N^M, and it follows that $h^D(e^D)$ is increasing in e^D.

To show that $h^D(e^D)$ is concave, consider the second derivative of $h^D(e^D)$:[21]

$$\begin{aligned}\frac{d^2h^D(e^D)}{de^D} = \frac{2v^2}{c^2}\cdot(Q_{in}^2-\frac{1}{4})\sum_{n^D=1}^{N}\binom{N}{n^D}p^{n_D}(1-p)^{N-n^D}\\ \sum_{n^M=1}^{N}\binom{N}{n^M}p^{N-n_M}(1-p)^{n^M}\cdot\frac{n^Dn^M(n^M-n^D)}{(n^D\cdot e^D+n^M\cdot e^M)^3}.\end{aligned}$$

We can split up the sum and consider the cases $n^M > n^D$ and $n^D \geq n^M$ separately. To prove that $h^D(e^D)$ is concave, it thus suffices to show that

$$\begin{aligned}\sum_{n^D=1}^{N}\binom{N}{n^D}p^{n_D}(1-p)^{N-n^D}\sum_{n^M=n^D+1}^{N}\binom{N}{n^M}p^{N-n_M}(1-p)^{n^M}\cdot\frac{n^Dn^M(n^M-n^D)}{(n^D\cdot e^D+n^M\cdot e^M)^3}-\\ \sum_{n^M=1}^{N}\binom{N}{n^M}p^{N-n_M}(1-p)^{n^M}\sum_{n^D=n^M}^{N}\binom{N}{n^D}p^{n_D}(1-p)^{N-n^D}\cdot\frac{n^Dn^M(n^D-n^M)}{(n^D\cdot e^D+n^M\cdot e^M)^3}\leq 0.\end{aligned}$$

We can rewrite this condition as

$$\begin{aligned}\sum_{n^D=1}^{N}\sum_{n^M=n^D+1}^{N}\binom{N}{n^D}\binom{N}{n^M}\cdot\frac{n^Dn^M(n^M-n^D)}{(n^D\cdot e^D+n^M\cdot e^M)^3}\cdot\Big[p^{n_D}(1-p)^{N-n^D}p^{N-n_M}(1-p)^{n^M}-\\ (1-p)^{n_D}p^{N-n^D}(1-p)^{N-n_M}p^{n^M}\Big]\leq 0.\end{aligned}$$

But this is equivalent to the inequality

$$\sum_{n^D=1}^{N}\sum_{n^M=n^D+1}^{N}\binom{N}{n^D}\binom{N}{n^M}\frac{n^Dn^M(n^M-n^D)}{(n^D\cdot e^D+n^M\cdot e^M)^3}\cdot\left[1-\left(\frac{p}{1-p}\right)^{2n^M-2n^D}\right]\leq 0,$$

and this clearly holds, since $p > p_0 > \frac{1}{2}$ and $n^M > n^D$ for all terms in the sum.

It remains to make the connection between the effort level e^D of the dominant group and the effort level e^* of the players with the group-preferred project. By definition, the two are equal (see the proof of Lemma 4.1). For example, A-players with project a are the dominant group at project a, but they are also the players with the group-preferred project among the players from group A. Similarly, the effort level e^M of the minority group and the effort level e^- of the players with the non-group preferred project are equal. For example, A-players with project b form the minority group at project b, and are the A-players that have chosen the non-group preferred project among A-players. □

[21] As before, we can ignore the case $n^D = n^M = 0$; and if $n^D = 0$ and $n^M > 0$, then the contribution to the sum is 0, and likewise for $n^D > 0, n^M = 0$.

D.7 Proof of Proposition 4.3

Recall the notation introduced in the proof of Lemma 4.1. By the results of Bollobás et al. (2007, p. 8, p. 10), the total number $N^D + N^M$ of players with a given project converges in probability to $|N|$, and the (random) fraction $\frac{N^D}{|N|}$ converges in probability to p. It is then straightforward to show that the fraction $\frac{N^D}{N^D+N^M}$ converges in probability to p. Hence, the function $h^D(e^D)$ (defined in the proof of Proposition 4.2) converges (pointwise) to

$$h^D(e^D) = \left(\frac{v}{c}\right) \cdot \left[\frac{p \cdot e^D \cdot Q_{in} + (1-p) \cdot e^M \cdot \frac{1}{2}}{p \cdot e^D + (1-p) \cdot e^M}\right].$$

The effort in an introspective equilibrium thus satisfies the fixed-point condition $e^D = h^D(e^D)$. This gives a quadratic expression (in e^D), which has two (real) solutions. One root is negative, so that this cannot be an introspective equilibrium by the proof of Proposition 4.2. The other root is as given in the proposition (where we have substituted e^D for e^*, e^M for e^- (see the proof of Proposition 4.2), and where we have used that $h = p - \frac{1}{2}$). □

D.8 Proof of Proposition B.1

First note that at level 1, the fraction p_1 of players with marker a that belong to group A is $p_1 := H_\varepsilon(0) > \frac{1}{2}$.

For $k > 1$, suppose that at level $k-1$, the fraction of players with marker a to group A is $p_{k-1} > \frac{1}{2}$. Moreover, suppose that each player j accepts proposals from anyone, and proposes only to players with the marker that is the group-preferred marker for player j's group. Then, at level k, an A-player chooses marker a if and only if

$$\tfrac{1}{2} \cdot \left(\Big[p_{k-1} \cdot Q_{in} + Q_{out} \cdot (1-p_{k-1})\Big] \cdot v + w_j^{A,a}\right) + \tfrac{1}{2} \cdot \left(Q_{in} \cdot v + w_j^{A,a}\right) \geq \\ \tfrac{1}{2} \cdot \left(\Big[p_{k-1} \cdot Q_{in} + Q_{out} \cdot (1-p_{k-1})\Big] \cdot v + w_j^{A,b}\right) + \tfrac{1}{2} \cdot \left(Q_{out} \cdot v + w_j^{A,b}\right).$$

The first term on the left- and right-hand side are the expected payoff if the player is the proposer (which happens with probability $\frac{1}{2}$). If an A-player is the proposer, he proposes to players with the group-preferred marker a, and interacts with a player from A with probability p_{k-1}, regardless of what marker he chose. If he is the responder, he gets proposals only from players for whom his marker is their group-preferred one (i.e., from A-players if he chose marker a; and from B-players if he chose marker b). So, at level k, the fraction p_k of players with marker a that belong to A is $p_k = H_\varepsilon(-\frac{1}{2} \cdot \beta)$, independent of k. It follows that the limiting fraction p of players with marker a that belong to A is

$$p = H_\varepsilon(-\tfrac{1}{2} \cdot \beta).$$

The result now follows from the definition of the tail distribution $H_\varepsilon(\cdot)$ (Appendix C.2). □

References

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

Alesina, A. and E. La Ferrara (2000). Participation in heterogeneous communities. *Quarterly Journal of Economics 115*, 847–904.

Alger, I. and J. W. Weibull (2013). Homo Moralis: Preference evolution under incomplete information and assortative matching. *Econometrica 81*, 2269–2302.

Alós-Ferrer, C. (1999). Dynamical systems with a continuum of randomly matched agents. *Journal of Economic Theory 86*, 245–267.

Aumann, R. J. (1987). Correlated equilibria as an expression of Bayesian rationality. *Econometrica 55*, 1–18.

Baccara, M. and L. Yariv (2013). Homophily in peer groups. *American Economic Journal: Microeconomics 5*, 69–96.

Baccara, M. and L. Yariv (2016). Choosing peers: Homophily and polarization in groups. *Journal of Economic Theory 165*, 152–178.

Bacharach, M. (1993). Variable universe games. In K. Binmore, A. Kirman, and P. Tani (Eds.), *Frontiers of Game Theory*. MIT Press.

Bardsley, N., J. Mehta, C. Starmer, and R. Sugden (2009). Explaining focal points: Cognitive hierarchy theory versus team reasoning. *Economic Journal 120*, 40–79.

Barth, F. (1969). Introduction. In F. Barth (Ed.), *Ethnic groups and boundaries*. Boston: Little, Brown.

Bénabou, R. (1993). Workings of a city: Location, education, and production. *Quarterly Journal of Economics 108*, 619–652.

Benhabib, J., A. Bisin, and M. O. Jackson (Eds.) (2010). *The Handbook of Social Economics*. Elsevier.

Bollobás, B., S. Janson, and O. Riordan (2007). The phase transition in inhomogeneous random graphs. *Random Structures & Algorithms 31*, 3–122.

Borgatti, S. P. and P. C. Foster (2003). The network paradigm in organizational research: A review and typology. *Journal of Management 29*(6), 991–1013.

Bramoullé, Y., S. Currarini, M. O. Jackson, P. Pin, and B. W. Rogers (2012). Homophily and long-run integration in social networks. *Journal of Economic Theory 147*, 1754–1786.

Cabrales, A., A. Calvó-Armengol, and Y. Zenou (2011). Social interactions and spillovers. *Games and Economic Behavior 72*, 339–360.

Calvó-Armengol, A., E. Patacchini, and Y. Zenou (2009). Peer effects and social networks in education. *Review of Economic Studies 76*, 1239–1267.

Camerer, C. F., T.-H. Ho, and J.-K. Chong (2004). A cognitive hierarchy model of games. *Quarterly Journal of Economics 119*, 861–898.

Camerer, C. F. and M. Knez (2002). Coordination in organizations: A game-theoretic perspective. In Z. Shapira (Ed.), *Organizational Decision Making*, Chapter 8. Cambridge University Press.

Camerer, C. F. and A. Vepsailanen (1988). The economic efficiency of corporate culture. *Strategic Management Journal 9*, 115–126.

Camerer, C. F. and R. A. Weber (2013). Experimental organizational economics. In R. Gibbons and J. Roberts (Eds.), *Handbook of Organizational Economics*, Chapter 6. Princeton University Press.

Charness, G., L. Rigotti, and A. Rustichini (2007). Individual behavior and group membership. *American Economic Review 97*, 1340–1352.

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

Costa-Gomes, M., V. P. Crawford, and B. Broseta (2001). Cognition and behavior in normal-form games: An experimental study. *Econometrica 69*, 1193–1235.

Costa-Gomes, M. A. and V. P. Crawford (2006). Cognition and behavior in two-person guessing games: An experimental study. *American Economic Review 96*, 1737–1768.

Craik, K. J. W. (1943). *The Nature of Explanation.* Cambridge University Press.

Crawford, V. P., M. A. Costa-Gomes, and N. Iriberri (2013). Structural models of nonequilibrium strategic thinking: Theory, evidence, and applications. *Journal of Economic Literature 51*, 5–62.

Crémer, J. (1993). Corporate culture and shared knowledge. *Industrial and Corporate Change 2*, 351–386.

Croson, R. T. A., M. B. Marks, and J. Snyder (2008). Groups work for women: Gender and group identity in the provision of public goods. *Negotiation Journal 24*, 411–427.

Currarini, S., M. O. Jackson, and P. Pin (2009). An economic model of friendship: Homophily, minorities, and segregation. *Econometrica 77*, 1003–1045.

Currarini, S. and F. Mengel (2016). Identity, homophily, and in-group bias. *European Economic Review*. Forthcoming.

de Vignemont, F. and T. Singer (2006). The empathic brain: How, when and why? *Trends in Cognitive Sciences 10*(10), 435–441.

Denzau, A. T. and D. C. North (1994). Shared mental models: Ideologies and institutions. *Kyklos 47*, 3–31.

DiMaggio, P. (1997). Culture and cognition. *Annual Review of Sociology 23*, 263–287.

Elfenbein, H. A. and N. Ambady (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin 128*, 243–249.

Epley, N. and A. Waytz (2010). Mind perception. In S. T. Fiske, D. T. Gilbert, and G. Lindzey (Eds.), *Handbook of Social Psychology* (Fifth ed.), Volume 1, Chapter 14. Wiley.

Galeotti, A. and L. Merlino (2014). Endogeneous job contact networks. *International Economic Review*. Forthcoming.

Golub, B. and M. O. Jackson (2012). How homophily affects the speed of learning and best-response dynamics. *Quarterly Journal of Economics 127*, 1287–1338.

Greif, A. (1994). Cultural beliefs and the organization of society: A historical and theoretical reflection on collectivist and individualist societies. *Journal of Political Economy 102*, 912–950.

Gruenfeld, D. H. and L. Z. Tiedens (2010). Organizational preferences and their consequences. In S. T. Fiske, D. T. Gilbert, and G. Lindzey (Eds.), *Handbook of Social Psychology* (Fifth ed.), Volume II, Chapter 33. John Wiley & Sons.

Hume, D. (1740/1978). *A Treatise of Human Nature* (Second ed.). Clarendon Press.

Jackson, M. O. (2008). *Social and Economic Networks.* Princeton University Press.

Jackson, M. O. (2014). Networks in the understanding of economic behaviors. *Journal of Economic Perspectives 28*, 3–22.

Jackson, M. O. and A. Wolinsky (1996). A strategic model of social and economic networks. *Journal of Economic Theory 71*, 44 – 74.

Jackson, M. O. and Y. Xing (2014). Culture-dependent strategies in coordination games. *Proceedings of the National Academy of Sciences 111*, 10889–10896.

James, W. (1890/1983). *Principles of Psychology.* Harvard University Press.

Kahneman, D. (2011). *Thinking, Fast and Slow.* Macmillan.

Kets, W. and A. Sandroni (2015). Challenging conformity: A case for diversity. Working paper.

Kossinets, G. and D. J. Watts (2009). Origins of homophily in an evolving social network. *American Journal of Sociology 115*, 405–450.

Kreps, D. M. (1990). Corporate culture and economic theory. In J. Alt and K. Shepsle (Eds.), *Perspectives on Positive Political Economy*, pp. 90–143. Cambridge University Press.

Kuran, T. and W. Sandholm (2008). Cultural integration and its discontents. *Review of Economic Studies 75*, 201–228.

Le Coq, C., J. Tremewan, and A. K. Wagner (2015). On the effects of group identity in strategic environments. *European Economic Review 76*, 239–252.

Lewis, D. (1969). *Convention: A Philosophical Study.* Cambridge, MA: Harvard University Press.

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

Mehta, J., C. Starmer, and R. Sugden (1994). The nature of salience: An experimental investigation of pure coordination games. *American Economic Review 84*, 658–673.

Mutz, D. C. (2002). Cross-cutting social networks: Testing democratic theory in practice. *American Political Science Review 0*, 111–126.

Nagel, R. (1995). Unraveling in guessing games: An experimental study. *American Economic Review 85*, 1313–1326.

Page, S. E. (2007). *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies*. Princeton University Press.

Patacchini, E. and Y. Zenou (2012). Ethnic networks and employment outcomes. *Regional Science and Urban Economics 42*, 938–949.

Pęski, M. (2008). Complementarities, group formation and preferences for similarity. Working paper, University of Toronto.

Pęski, M. and B. Szentes (2013). Spontaneous discrimination. *American Economic Review 103*, 2412–2436.

Robalino, N. and A. J. Robson (2015). The evolution of strategic sophistication. *American Economic Review*. forthcoming.

Schelling, T. (1960). *The Strategy of Conflict*. Harvard University Press.

Sethi, R. and R. Somanathan (2004). Inequality and segregation. *Journal of Political Economy 112*, 1296–1321.

Skyrms, B. (1996). *Evolution of the Social Contract*. Cambridge University Press.

Stahl, D. O. and P. W. Wilson (1995). On players' models of other players: Theory and experimental evidence. *Games and Economic Behavior 10*, 218–254.

Sugden, R. (1995). A theory of focal points. *Economic Journal 105*, 533–550.

Sugden, R. (1998). The role of inductive reasoning in the evolution of conventions. *Law and Philosophy 17*, 377–410.

Van den Steen, E. (2010). Culture clash: The costs and benefits of homogeneity. *Management Science 56*, 1718–1738.

Weber, R. A. (2006). Managing growth to achieve efficient coordination in large groups. *American Economic Review 96*, 114–126.

Weber, R. A. and C. F. Camerer (2003). Cultural conflict and merger failure: An experimental approach. *Management Science 49*, 400–415.