A belief-based theory of homophily*

Willemien Kets† Alvaro Sandroni‡

December 31, 2014

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 in each other's shoes when they share a similar background. The model delivers novel comparative statics that emphasize the interplay of cultural and economic factors. We investigate how the socially optimal level of homophily varies with the economic environment.

*We thank Sandeep Baliga, Vincent Crawford, Georgy Egorov, Tim Feddersen, Matthew Jackson, Rachel Kranton, Nicola Persico, Yuval Salant, Rajiv Sethi, Eran Shmaya, Andy Skrzypacz, Jakub Steiner, and numerous seminar audiences and conference participants for helpful comments and stimulating discussions.

†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 (Benhabib et al., 2010), to organizational research (Borgatti and Foster, 2003), social psychology (Gruenfeld and Tiedens, 2010), political science (Mutz, 2002), and sociology (McPherson et al., 2001). Homophily produces segregated social and professional networks, affect hiring and promotion decisions, investment in education, and the diffusion of information. For decades, these matters have been at the center of political confrontations and public policy. 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, 2008, for a survey). However, without a theory of the determinants of these preferences, it is difficult to explain why homophily is observed in some cases, but not in others (beyond positing homophilous preferences only in the former cases). Moreover, it is difficult to assess the welfare implications of policy interventions in such models.

We provide a theory of homophily that is not based on homophilous preferences. Rather, we explain these preferences by showing that people can gain by interacting with similar others if that reduces strategic uncertainty. This makes it possible to evaluate different policy interventions. Moreover, it allows us to derive clear and intuitive comparative statics.

Our starting point is that to understand people's tendency to interact with similar others, we need to unpack the black box of cultural identity. Following Kreps (1990), we view culture as a means to reduce strategic uncertainty. When there is uncertainty about what action is appropriate, cultural rules can act as focal principles. As a metaphor for situations that are characterized by strategic uncertainty, we consider (pure) coordination games like the following:

	s^1	s^2
s^1	1,1	0,0
s^2	0,0	1,1

Such games are rife with strategic uncertainty: the payoff structure provides little guidance. However, there may be a focal point, which may depend on the context in which the game is played. If players share a cultural code, and may successfully coordinate by jointly inferring the focal point from the context of the game.

To model players' behavior, we build on the dual process account of Theory of Mind in psychology.[1] The dual process account of Theory of Mind posits that an individual can form

[1]See Epley and Waytz (2010) for a survey. The dual process account of Theory of Mind relies on a rapid

an instinctive understanding of another person's beliefs, by putting himself in the other's position, and then adapt his views through reasoning. To capture this, we assume that each player has some initial (random) impulse telling him which action is appropriate. A player's first reaction is to follow his impulse. After some introspection, the player realizes that he may act on instinct and so, his opponent may also act on instinct. In addition, the impulse of the opponent may be similar to his own. This means that impulses can be used as input to form initial beliefs. But if the player thinks a little more, he realizes that his opponent may have gone through a similar reasoning process, leading the player to revise his initial beliefs. This process continues to higher orders. The limit of this procedure, where players go through the entire reasoning process in their mind before making a decision, defines an *introspective equilibrium*.

A key assumption is that players find it easier to project themselves into the role of players that are similar to them, in line with neurological and behavioral evidence (de Vignemont and Singer, 2006; Jackson and Xing, 2014). In the context of our model, this is because players with a shared cultural background may have similar ideas about which action is appropriate in a given context. Formally, players belong to different groups, and initial impulses are (imperfectly) correlated within groups and independent across groups. So, if impulses are highly correlated, players from the same group are likely to agree on which action is focal in a particular context.

Our first result shows that in the unique introspective equilibrium, each player follows his initial impulse. So, the naive response of following one's initial impulse is optimal under the infinite process of reasoning through higher-order beliefs. This holds even if impulses are noisy and people from different subcultures have uncorrelated impulses. It now follows that players coordinate more effectively if they belong to same group. This provides an incentive to choose activities that enhance the chances of meeting similar people, that is, to be *homophilous*.[2]

If group membership is not observable, players can enhance their chances of interacting with members of their own group, for example, by choosing a project (e.g., a hobby, profession, or neighborhood). We consider an an extended game where players first choose a project and subsequently play the coordination game with an opponent that has chosen the same project. We analyze this extended game using the same method as before. Players introspect on their impulses, use them to form initial beliefs and modify them through higher-order reasoning.

instinctive process and a slower cognitive process. As such, it is structurally similar to the two-systems account of decision-making under uncertainty, as popularized by Kahneman (2011). The foundations of dual process theory go back to the work of the psychologist William James (1890/1983). See Section 2 for an extended discussion.

[2]Alternatively, players could reduce the risk of miscoordinating by learning the cultural code of the other group (Lazear, 1999). However, this may be costly.

We show that there is a unique introspective equilibrium. In the unique equilibrium, similar players overwhelmingly choose the same project, regardless of their intrinsic sentiments over projects. Thus, the introspective process leads players to coordinate effectively. The level of homophily depends on economic incentives (i.e., the coordination payoffs) and the strength of cultural identity (i.e., similarity in impulses within a group). So, when focal points depend on cultural background and on the context, introspection leads to homophily.

These results allow us to explain why we observe homophily in some settings, but not in others, even if the underlying game is the same. For example, in rapidly changing environments, organizational culture may not strongly guide initial impulses, and this leads to more integrated networks (Staber, 2001). Consistent with these observations, our model predicts that the level of homophily is lower when cultural identity is weaker. In addition, the mechanism that leads people to seek out similar others may also lead them to become similar on other dimensions as well. For example, in order to associate with others that have similar values, people choose the same hobbies, professions, or clubs as they do (Kossinets and Watts, 2009).[3]

We then turn to the question of optimal social structure. We consider a policy maker who faces uncertainty about the game. With some probability, both actions are equally profitable; and with some probability, coordination on one of the actions is more profitable than coordination on the other, and potentially much more so, perhaps following an innovation. The policy maker chooses the level of homophily before payoffs are realized. Players observe the payoffs and then play the game, following the introspective process introduced earlier. We investigate how the social structure that maximizes social welfare depends on the economic environment.

If the economic environment is stable, so that with high probability, the game is close to the pure coordination game considered above, it is socially optimal to have high levels of homophily. Intuitively, increasing the probability that players interact with members of their own group reduces strategic uncertainty. Naturally, strengthening players' cultural identity further reduces strategic uncertainty, and increases social welfare, even if it leads to segregation. This could shed light on the effects of policies aimed at strengthening cultural identity, for example in organizations (e.g., Tichy and Sherman, 2005) and at the national level (Macdonald, 2012).

The conclusions are starkly different in an uncertain economic environment, where there is a high probability that an innovation makes one of the actions much more profitable than the others. In this case, social welfare is maximal if there is a minority of a significant size.

[3]This is different from the well-known phenomenon that individuals who interact frequently influence each other, and thus become more similar in terms of behavior (e.g., Benhabib et al., 2010). Here, becoming more similar is a pre-condition for interaction, not the result thereof.

Intuitively, a member of the minority group faces less pressure to conform than a member of the majority group. This is because a player from the minority group is unlikely to be matched with a player from his own group, and thus faces only a weak incentive to select the action that he thinks members of his own group are likely to take. He may thus take the high-payoff action even if it goes against his impulse. To maximize social welfare, this non-conformist behavior of the minority has to spill over to the majority, so that all players choose the high-payoff action. This can happen only if the minority has a critical mass. Again, strengthening players' cultural identity reduces strategic uncertainty, but the welfare effects are now opposite: social welfare is lower when cultural identity is stronger. Intuitively, the stronger players' cultural identity, the stronger the pressure to conform, and thus the harder it is for players to choose the high-payoff action when it is against their impulse.

So, whereas a segregated society with a strong cultural identity does well in a stable economic environment, the same factors that make such a society successful at coordinating in a stable economic environment limits its flexibility when an innovation may make one of the actions superior. By contrast, more integrated societies with a weaker cultural identity have the agility to perform well in such an uncertain environment. Formally modeling a mechanism that can capture these common intuitions has several benefits. First, it allows us to characterize the conditions under which segregation or integration are socially optimal. Moreover, it allows us to ask how the optimal size of the minority varies with primitives such as the strength of players' cultural identity and the coordination payoff. Finally, it suggests a rationale for promoting diversity even in the absence of complementarity of skills across groups: a more diverse society harbors dissent, and dissent may prevent excessive conformism in the face of innovations, even if it hampers players' ability to coordinate in other settings.

Thus far, we have assumed that all players are matched with exactly one other player. Section 5 develops an extension of the model that can capture network formation, by allowing players to interact with multiple players at a cost. The level of homophily can now be even higher, as greater success in coordinating with similar others translates into greater incentives to form connections. In addition, the model accommodates important properties of social and economic networks. Since these features arise endogenously, the model provides novel testable hypotheses about how these properties change when the fundamentals vary. For example, when cultural identity is strong, networks tend to be densely connected, with high levels of homophily and significant inequality in the number of connections. This means that the network consists of a tightly connected core of gregarious players from one group, with a periphery of hermits from the other group that are loosely connected with the core, consistent with empirical observations of social and economic networks (Jackson, 2008).

The heart of our contribution lies in the modeling of players' reasoning process. In a setting where standard equilibrium refinements have no bite,[4] we show that simply taking into account players' reasoning process is a powerful method to obtain uniqueness in a range of different settings. This equilibrium uniqueness in turn allows us to derive intuitive yet subtle comparative statics, and allows us to assess the welfare implications of commonly used policies.

1.1. Related literature

Homophily is a widespread phenomenon that has important economic implications, affecting hiring and promotion decisions, the spread of information, and educational outcomes (Jackson, 2014). The literature on homophily in economics mostly assumes homophilous preferences and investigates the implications for network structure and economic outcomes (e.g., Schelling, 1971; Alesina and La Ferrara, 2000; Currarini et al., 2009; Golub and Jackson, 2012; Alger and Weibull, 2013), with Baccara and Yariv (2013) and Peski (2008) being notable exceptions.[5] By contrast, we derive players' incentives to interact with similar others from a desire to reduce strategic uncertainty. This makes it possible to obtain intuitive comparative statics and to evaluate the tradeoffs inherent in diversity policies. In its aim to explain interaction patterns from underlying economic drivers, our paper is related on the literature on residential segregation (e.g., Bénabou, 1993, 1996; Durlauf, 1996; Cutler and Glaeser, 1997). Unlike much of that literature, we abstract away from specific sources of complementarities or externalities. Our results thus suggest that segregation can be a natural outcome in a wide range of environments, even if none of the mechanisms previously considered in the literature (e.g., peer effects, externalities in public good provision) are present. Moreover, in contrast with much of the literature, there is no asymmetry across groups in skills, wealth, or spillovers in our model (either exogenous or endogenously derived). Again, this greatly expands the range of settings where homophily and segregation can be expected to arise, and it allows us to analyze measures aimed at influencing interactions when no group is universally seen as a

[4]Like other models that explain homophily and segregation, the games we study have multiple equilibria with sometimes very different properties (Appendix B). While 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), such refinements are not always strong enough to give uniqueness.

[5]In a public good provision model, Baccara and Yariv (2013) show that groups are stable if and only if their members have similar preferences. Peski (2008) shows that segregation is possible if players have preferences over the interactions that their opponents have with other players (also see Peski and Szentes, 2013). No such assumption is needed for our results. Also, Greif (1993) shows that it may be optimal for individuals to interact with members of the own group if there are market imperfections. We show that homophily may be optimal even in the absence of market imperfections.

more desirable partner than another.

The process we consider bears some resemblance with level-k models (see Crawford et al., 2013, for a survey). A key difference is that we are interested in equilibrium selection, while the level-k literature focuses on nonequilibrium behavior. Indeed, one contribution of our paper is that it demonstrates that modeling players' reasoning process can give rise to new insights even if one focuses on the equilibrium limit. Another difference is that the level-k literature does not consider payoff-irrelevant signals such as impulses, which are critical in our setting. Our model is also very different from global games (e.g., Morris and Shin, 2003), as there is no payoff uncertainty in our model. Importantly, global games do not select an equilibrium in all pure coordination games, while our process does.[6]

Our work sheds light on experimental findings that social norms and group identity can lead to an efficient equilibrium and can improve coordination, as in the minimum-effort game (Weber, 2006; Chen and Chen, 2011), the provision point mechanism (Croson et al., 2008) and the Battle of the Sexes (Charness et al., 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 with a similar background. Our mechanism operates even if no equilibrium is superior to another, as in some pure coordination games.

2. Coordination, culture and introspection

There are two groups, A and B, each consisting of a unit mass of players. Members of these groups are sometimes called A-players and B-players, respectively. Group membership is not observable. Players are matched with an opponent of the same 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 that are matched interact in a coordination game, with payoffs given by:

	s^1	s^2
s^1	v,v	0,0
s^2	0,0	v,v

Payoffs are commonly known. Nature draws a (payoff-irrelevant) state $\theta_G = 1, 2$ for each group $G = A, B$, independently across groups. The state is the focal point of the group. So,

[6]The introspective process also bears some formal resemblance to the deliberative process introduced by Skyrms (1990). Skyrms focuses on the philosophical underpinnings of learning processes and the relation with classical game theory.

if $\theta_A = 1$ then the culture of A-players takes s^1 to be the appropriate action in the current context. Ex ante, states 1 and 2 are equally likely for both groups.

Each player has an initial impulse to take an action. Their impulse is influenced by their culture. That is, a player's initial impulse is more likely to match the focal point of his group than the alternative action. So, if $\theta_A = 1$ then A-players initial impulse is to take action s^1 with probability $q > \frac{1}{2}$, independently across players. The same statement holds for B-players. When q is close to 1, a player's culture strongly guides initial impulses. When q is close to $\frac{1}{2}$, a player's culture has a negligible influence on initial impulses. Thus, players have an imperfect understanding of their cultural code.

A player's first instinct is to follow his initial impulse, without any strategic considerations. We refer to this initial stage as level 0. At higher levels, players realize that if their opponent is in the same group, then they are likely to have a similar impulse. So, by introspecting (i.e., by observing their own impulses), players obtain an informative signal about what their opponents will do. At level 1, a player formulates a best response to the belief that his opponent will follow her impulse. This process needs not stop at level 1. At level $k > 1$, players formulate a best response to the belief that the opponent is at level $k - 1$. Together, this constitutes a reasoning process of increasing levels. These levels do not represent actual behavior; they are constructs in a player's mind. We are interested in the limit of this process, as the level k goes to infinity. If such a limit exists for each player, then the profile of such limiting strategies is referred to as an *introspective equilibrium*.

This approach is motivated by the dual process account of the Theory of Mind in psychology (Apperly, 2012; Baron-Cohen et al., 2013; Epley and Waytz, 2010; Fiske and Taylor, 2013). The key idea behind this approach is that reasoning about other people's beliefs and desires involves reasoning about unobservable mental states, which starts from a base of readily accessible knowledge and proceeds by adjusting instinctive responses in light of less accessible information, for example, how the other person's mental state may differ from one's own. So, while people have instinctive reactions (modeled here with impulses), they may modify their initial views using theoretical inferences about others (captured here by the different levels).[7,8]

[7]These ideas have a long history in philosophy. According to Locke (1690/1975) people have a faculty of "Perception of the Operation of our own Mind" which, "though it be not Sense, as having nothing to do with external Objects; yet it is very like it, and might properly enough be call'd internal Sense," and Mill (1872/1974) writes that understanding others' mental states first requires understanding "my own case." Kant (1781/1997) suggests that people can use this "inner sense" to learn about mental aspects of themselves, and Russell (1948) observes that "[t]he behavior of other people is in many ways analogous to our own, and we suppose that it must have analogous causes."

[8]Kimbrough et al. (2013) interpret Theory of Mind as the ability to learn other players' payoffs, and shows that this confers an evolutionary benefit in volatile environments.

A critical assumption is that players' impulses are correlated (perhaps mildly) within groups (i.e., $q > \frac{1}{2}$), that is, a player's own impulses are informative of the impulses of similar players. Thus, players find it easier to put themselves in the shoes of those from their own group. This is consistent with experimental evidence from neuroscience and psychology that shows that it is easier to predict the behavior or feelings of similar people (de Vignemont and Singer, 2006). This is also supported by experimental studies in economics (Currarini and Mengel, 2013; Jackson and Xing, 2014).

Our first result shows that the seemingly naive strategy of following one's initial impulse is the optimal strategy that follows from the infinite process of high order reasoning.

Proposition 2.1. *There is a unique introspective equilibrium. In this equilibrium, each player follows his initial impulse.*

So, 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 his impulse, it is optimal for her to do so as well; this, in turn, makes it optimal for the opponent to follow his impulse.

As is well-known, coordination games have multiple (correlated) equilibria. For example, all players choosing action s^1, regardless of their signal, is a correlated equilibrium. Moreover, standard equilibrium refinements have no bite in pure coordination games such as the one considered here. By contrast, the introspective process selects only one equilibrium. This uniqueness will prove critical for predictions and comparative static results in the next sections.[9]

While it is natural to assume that nonstrategic players follow their impulse, our results do not depend on this. For example, as long as each player is more likely than not to follow his impulse, our results continue to hold. What is needed for our results is that players do not have a strong predisposition to choose a fixed action, independent of context. Also, the result does not hinge on the states of the groups being independent, or on impulses coming in the form of action recommendations (as opposed to, say, beliefs about the other's belief or action).

Let $Q := q^2 + (1-q)^2 > \frac{1}{2}$ be the odds that two players from the same group have the same initial impulse. If Q is close to 1, impulses are strongly correlated within a group. If Q is close to $\frac{1}{2}$, impulses within a group are close to independent, as they are across groups. We refer to Q as the strength of players' *cultural identity*. Indeed, in a complex and unpredictable world,

[9] Bacharach and Stahl (2000) similarly show that if nonstrategic players favor a certain option in a coordination game, then this advantage gets magnified at higher levels. However, they focus on nonequilibrium outcomes, and their procedure does not guarantee uniqueness.

cultural identity is a critical means to simplify otherwise excessive information flows (Jenkins, 2014, Ch. 12). While we focus on this particular aspect of identity, our notion of cultural identity is broad in its scope: it encompasses social, ethnic, religious, and organizational identity, among others, as these can all be a source of greater predictability.

In the unique introspective equilibrium, expected payoffs are:

$$\left[pQ + (1-p) \cdot \tfrac{1}{2}\right] \cdot v.$$

Thus:

Corollary 2.2. *For every $Q > \frac{1}{2}$, the expected utility of a player strictly increases with the probability p of being matched with a player from the own group.*

Similar players are more likely to coordinate their actions on the focal point determined by their culture (which may be context-dependent). If players share common cultural identity, they are more likely to coordinate their actions on the focal point determined by their culture (which may be context-dependent). This is consistent with experimental evidence that shows that focal points may differ across groups, and may depend on the fine details of the decision context (Weber and Camerer, 2003; Bardsley et al., 2009).

By Corollary 2.2, players have an incentive to seek out similar players,[10] consistent with work in social psychology and sociology showing that people want to interact with members of their own group to reduce uncertainty (Hogg, 2007; Jenkins, 2014). 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 to prefer project a. Specifically, for A-players, the value of project a is drawn uniformly at random from $[0, 1]$, while the value of project b is drawn uniformly at random

[10] Similar results have been shown in other settings. See, for example, Phelps (1972) and Cornell and Welch (1996) on hiring practices, Sethi and Yildiz (2014) on prediction and information aggregation, and Crawford (2007) and Ellingsen and Östling (2010) on coordination and communication.

from $[0, 1-2\varepsilon]$, for small $\varepsilon > 0$. For B-players, an analogous statement holds with the roles of projects a and b reversed. So, 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 (Appendix A). Thus, project a is the *group-preferred project* for group A, and project b is the group-preferred project for group B. Such a slight asymmetry in preferences between could result if some project fits better with culture-specific norms than others (Akerlof and Kranton, 2000). Players' payoffs are the sum of the intrinsic value of the chosen project and the (expected) payoff from the coordination game.

Players follow the same process as before. At level 0, players follow their impulse and select the project they intrinsically prefer. At level $k > 0$, players formulate a best response to actions selected at level $k-1$: a player chooses project a if and only if the expected payoff from 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. *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: $p^a = p^b$.*

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. Let the *level of homophily* $h := p - \frac{1}{2}$ 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. When the level of homophily is close to 0, there is almost full integration. When the level of homophily is close to $\frac{1}{2}$, there is nearly complete segregation.

There is a fundamental difference between exogenous and endogenous matching. When matching is exogenous, as in Section 2, players end up following their impulses after they have gone through the entire reasoning process. Thus, their ability to successfully coordinate is determined by their cultural identity, that is, by the degree to which their culture shapes their initial impulses over actions. In contrast, in the case of endogenous matching, players may not act on impulse. Intuitively, at level 1, player realize that there is a slightly higher chance of meeting a similar player if they choose the group-preferred project. So, players may select the group-preferred project even if their intrinsic value for the alternative project is slightly higher. At level 2 an even higher fraction of agents may select the group-preferred project because the odds of finding a similar player this way are now higher than at level 1. So, the attractiveness of the group-preferred project is reinforced throughout the entire process in this case. It is possible that all players, even those who have a strong intrinsic preference for the alternative project, choose the group-preferred project. Complete segregation may arise even in cases where there would be almost complete integration if players were to act on their initial impulses (i.e., ε small). In this sense, introspection and reasoning are root causes of segregation. This intuition is formalized in the next result.

Proposition 3.2. *There is a unique introspective equilibrium. In the unique equilibrium, there is complete segregation ($h = \frac{1}{2}$) if and only if*

$$v(Q - \tfrac{1}{2}) \geq 1 - 2\varepsilon.$$

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

$$h = \frac{(1-2\varepsilon)}{4v^2(Q-\frac{1}{2})^2} \cdot \left[2v(Q-\tfrac{1}{2}) - 1 + \sqrt{\frac{4v^2(Q-\frac{1}{2})^2}{1-2\varepsilon} - 4v(Q-\tfrac{1}{2}) + 1}\right].$$

In any case, the equilibrium level of homophily exceeds the initial level of homophily (i.e., $h > \varepsilon$).

Proposition 3.2 characterizes the introspective equilibrium. In the unique equilibrium, a large share of players choose the group-preferred project. In fact, strategic considerations *always* produce more segregation than would follow from differences in intrinsic preferences over projects alone (i.e., $h > \varepsilon$). The result demonstrates that a strong cultural identity may give rise to segregation. If cultural identity is sufficiently strong, then all players choose the group-preferred project, regardless of their intrinsic preferences. So, introspection and reasoning may lead to complete segregation even if players do not have any direct preferences for interacting with similar others and, ex ante, players have arbitrarily similar preferences over projects (i.e., ε small). The comparative statics for the level of homophily follow directly from Proposition 3.2:

Corollary 3.3. *The level of homophily h increases with the strength of the cultural identity Q and with the coordination payoff v. Cultural identity and economic incentives are complements: the level of homophily is high whenever either cultural identity or the coordination payoff is high.*

Figure 1 shows the level of homophily as a function of the coordination payoff v and the strength of players' cultural identity Q. Regardless of the strength of the cultural identity, the level of homophily increases with economic incentives to coordinate. These comparative statics results deliver clear and testable predictions for the model. That is, even if it is not

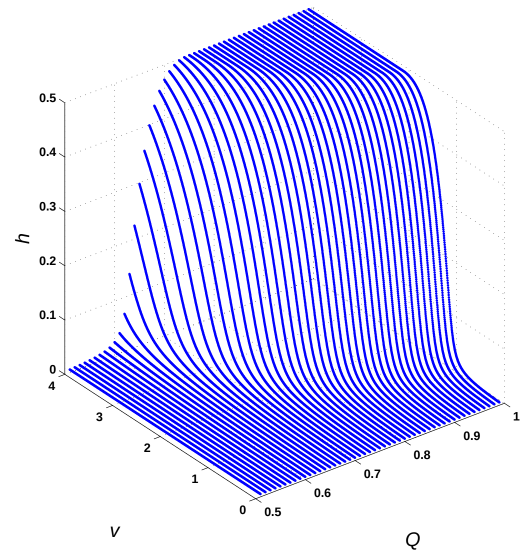

Figure 1: The equilibrium level of homophily h as a function of the coordination payoff v and the strength of players' cultural identity Q.

possible to observe the strength of players' cultural identity, the model still predicts a positive correlation between coordination payoffs and homophily. Also, when cultural rules provide clear guidance (i.e., Q close to 1), the level of homophily increases.

While intuitive, these predictions require a form of equilibrium selection, which we obtain here through the dual process account of Theory of Mind. Standard analysis delivers a multiplicity of equilibria. Some of these equilibria are highly inefficient. For example, there may be equilibria in which all players choose the non-group preferred project (e.g., all A-players choose project b); see Appendix B. In such equilibria, the majority of players choose a project that they do not intrinsically prefer. Choosing a project constitutes a coordination problem, and inefficient lock-in can occur in equilibrium. In contrast, when players are introspective, the majority always chooses the group-preferred project, society avoids inefficient lock-in, and successful coordination on the payoff-maximizing outcome ensues. In turn, this gives rise to unambiguous and intuitive comparative statics for the introspective equilibrium.

Our framework suggests that any aspect of identity that affects predictability, like religion, a shared upbringing, educational background or profession, may be a basis for homophily, while other aspects, such as height, are less likely sources. Thus, our framework captures what sociologists call value homophily (McPherson et al., 2001). By emphasizing predictability and strategic uncertainty, our model can shed light on why preferences for interacting with other groups are often situational. For example, homophily on the basis of race is reduced substantially when individuals are similar on some other dimension, such as socioeconomic status (Park et al., 2013). This suggest that individuals do not have some immutable preference or dislike of other groups. It can also help understand why individuals may have a tendency to identify with groups that strongly distinguish themselves in their values and practices, even if these distinctions are valued negatively (Ashford and Mael, 1989). Also, if homophily is driven by a desire to reduce strategic uncertainty, this may explain interaction patterns when identities are nested. As strategic uncertainty is reduced more when groups have a stronger identity, individuals may have a preference for interacting with people that match their narrowly defined identity rather than a more broadly defined one (e.g., Korean-Americans vs. Asian-Americans); but when given the choice between individuals that fit their broader identity and people with a distinct cultural background, they prefer to interact with someone with whom they share some some commonality (e.g., Asian-Americans vs. Americans at large), consistent with empirical evidence (Nagel, 1994; Ashford and Johnson, 2014). Finally, the incentives for segregation are not affected by the type of the other group in our model, provided that the degree of strategic uncertainty is kept constant. If 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 level of homophily remains unchanged. This is consistent with empirical evidence which shows that homophily often stems from beneficial interaction with similar players, rather than a dislike of a particular group of outsiders (e.g., Marsden, 1988; Jacquemet and Yannelis, 2012). While these features can potentially be captured by models that directly posit homophilous preferences (e.g., Alesina and La Ferrara, 2000), this would require tailoring preferences to observed phenomena. More fundamentally, it would not allow one to predict interaction patterns ex ante, from primitives.

Our results do not depend on our specific assumptions, such as the exact assumptions on preferences or the signal structure.[11] Moreover, similar results obtain in variations of the model. In Appendix C, we show that our results also go through if players cannot sort by choosing projects, but instead can signal their identity using markers, that is, observable attributes such as tattoos or specific attire. Again, high levels of homophily can arise in equilibrium, with a large share of players choosing the group-preferred marker. These results help explain why groups are often marked by seemingly arbitrary traits (Barth, 1969).

On the other hand, our framework can also be used to investigate the conditions under which homophily is limited. One possibility is that players from different groups can have complementary skills. Our model can easily accommodate this possibility, by assuming that

[11] For example, the assumption that there are group-preferred projects can be relaxed substantially. All we need is that there is some asymmetry in intrinsic preferences over projects between groups. In particular, our results go through if a (large) majority of both groups (intrinsically) prefer a certain project, say a, as long as one of the groups has an even stronger preference for that project than the other. 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 $\bar{u}$, independent of which other players choose this option or what further actions players take.

players receive a payoff $V > v$ if they coordinate with someone from the other group (and a payoff v if they coordinate with a member of their own group). This makes that players need to trade off the greater likelihood of successfully interacting with the own group with the higher payoffs from interacting with the other group, conditional on successful coordination. In Appendix D, we define the marginal benefit of interacting with the own group, which captures this tradeoff. We show that there is significant homophily in equilibrium if and only if the marginal benefit of interacting with the own group is positive.

4. Welfare and policy implications

Our model can help elucidate economic tradeoffs inherent in policies aimed at enhancing diversity or strengthening cultural identity. We consider a policy maker who aims to maximize social welfare, that is, the sum of coordination payoffs and product values. As social interaction patterns are typically stable and do not readily adjust, a policy maker may want to take into account how payoffs are likely to change in setting his policy. To capture this, we consider a policy maker that faces uncertainty about payoffs at the time he chooses a policy. Specifically, one of the actions may give higher payoffs than the other, perhaps following a technological innovation. That is, the game is given by:

	s^1	s^2
s^1	v^*,v^*	0,0
s^2	0,0	v,v

where $v^* \geq v > 0$. Thus, successful coordination on action s^2 gives players a payoff of v, as before, while coordinating on s^1 gives them a potentially higher payoff, v^*. We refer to action s^1 as the Pareto superior action. Before payoffs are realized, the policy maker chooses the level of homophily to maximize social welfare, that is, the sum of coordination payoffs and the value of derived from projects. After payoffs are realized, players observe the payoffs v and v^* to both actions and play the coordination game, taking the level of homophily as given. When choosing their action in the coordination game, they follow the same introspective process described earlier: each player has an impulse, and goes through infinitely many levels of introspection.[12]

We contrast the case where a technological innovation is unlikely to the case where the gains from an innovation can be large. We capture this by assuming that the ratio v^*/v of

[12] Again, the results presented in this section are robust to changes in distributional assumptions. In addition, similar results are obtained in a multi-period extension where payoffs are fixed for some time, with an innovation occurring at some random time.

coordination payoffs follows a Pareto distribution. Thus, the probability that the ratio v^*/v is at least $y \geq 1$ is $y^{-\alpha}$, for some $\alpha > 1$. When α is large, the payoff v^* equals v with high probability, and we say that the economic environment is stable. In the limit $\alpha = \infty$, we are back in the benchmark case where payoffs are fixed: $v^* = v$ with probability 1. On the other hand, if α is close to 1, then the expected value of v^* grows arbitrarily large, and we say that the economic environment is uncertain. The parameter α thus measures the degree of economic stability.

The *socially optimal level of homophily* h^α is the level of homophily that maximizes social welfare $W^\alpha(h)$, given by

$$W^\alpha(h) = C^\alpha(h) + \Pi(h),$$

where $C^\alpha(h)$ is the total coordination payoff and $\Pi(h)$ is the total value that players assign to projects (assuming that the share $p = h + \frac{1}{2}$ of players with the highest intrinsic preference for the group-preferred project chooses that project).[13]

4.1. Stable economic environments

In a stable economic environment, there is only a small probability that a technological innovation increases the payoffs of one of the actions. In the limit $\alpha = \infty$, payoffs are fixed: the coordination payoff to either action is v. This is the benchmark case we have studied so far. The next result shows that the socially optimal level of homophily in a stable economic environment is arbitrarily close to the socially optimal level of homophily when payoffs are fixed.

Proposition 4.1. *The socially optimal level of homophily in a stable economic environment is arbitrarily close to the socially optimal level of homophily when payoffs are fixed, that is, h^α converges to h^∞ as $\alpha \to \infty$. Moreover, social welfare also converges: $W^\alpha(h^\alpha) \to W^\infty(h^\infty)$.*

In the remainder of this section, we concentrate on the benchmark case, for simplicity. By Proposition 4.1, the results hold approximately for stable economic environments where there is a small probability of a technological innovation. The next result characterizes the socially optimal level of homophily for the benchmark case:

Proposition 4.2. *Suppose payoffs are fixed (i.e., $\alpha = \infty$). Full segregation is socially optimal (i.e., $h^\infty = \frac{1}{2}$) if and only if*

$$Q \cdot (v - \tfrac{1}{2}) \geq \tfrac{1}{2} - \varepsilon.$$

[13] We can thus think of the policy maker as setting a threshold t such that players whose intrinsic preference for the group-preferred project is at least t choose that project, in such a way that the share of the players choosing the group-preferred project equals $p = h + \frac{1}{2}$.

If full segregation is not socially optimal (i.e., $h^\infty < \frac{1}{2}$), then the socially optimal level of homophily is given by:

$$h^\infty = \frac{(1-2\varepsilon)}{4(Qv - \frac{1}{2}V)^2} \cdot \Big[Qv - \tfrac{1}{2}V - \frac{1}{4} + \sqrt{\frac{4(Qv-\frac{1}{2}V)^2}{1-2\varepsilon} - \tfrac{1}{2}(Qv - \tfrac{1}{2}V) + \tfrac{1}{16}}\Big].$$

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

When the level of homophily h is high, strategic uncertainty is limited, so that the total coordination payoff $C^\infty(h)$ is high. However, players may choose a project that they do not intrinsically prefer, and this is reflected in the value derived from projects $\Pi(h)$. The social optimum trades off these two factors.

Like the equilibrium level of homophily, the socially optimal level of homophily depends on the strength of players' cultural identity Q and on the economic benefits of coordination v. When players' cultural identity is strong and the economic benefits of coordination are high, there are large gains from coordination, and the socially optimal level of homophily is high. On the other hand, if cultural identity and economic incentives are weak, the benefit from players choosing their intrinsically preferred project is relatively large, and the socially optimal level is low. Thus:

Corollary 4.3. *Suppose payoffs are fixed (i.e., $\alpha = \infty$). The socially optimal level of homophily h^∞ increases with the strength of the cultural identity Q and with the coordination payoff v. Cultural identity and economic incentives are complements: the level of homophily is high whenever either cultural identity or the coordination payoff is high.*

As in the introspective equilibrium, the socially optimal level of equilibrium increases with economic incentives and the strength of players' cultural identity; and if cultural identity and economic incentives are sufficiently strong, then full segregation is socially optimal. This is illustrated in Figure 2.

We can compare the socially optimal level of homophily to the equilibrium level. It turns out that there can be too little homophily in equilibrium:

Corollary 4.4. *Suppose payoffs are fixed (i.e., $\alpha = \infty$). The level of homophily in the unique introspective equilibrium never exceeds the socially optimal level of homophily; and if $v \cdot (Q - \frac{1}{2}) \leq 1 - 2\varepsilon$, the equilibrium level of homophily is strictly below the socially optimal level of homophily.*

The equilibrium level of homophily can be substantially lower than in the social optimum. If cultural identity and economic benefits are of intermediate strength, full segregation is socially optimal, while there is only partial segregation in equilibrium. Intuitively, there are

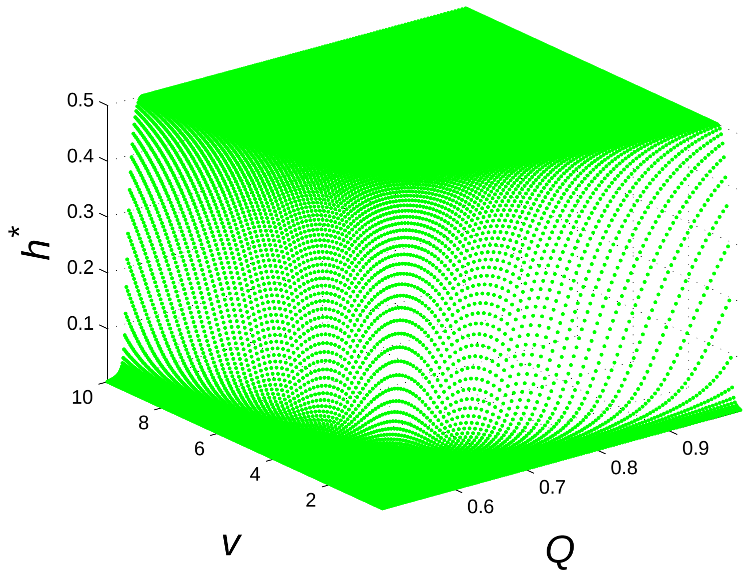

Figure 2: The socially optimal level of homophily as a function of the coordination payoff v and the strength of players' cultural identity Q when payoffs are fixed ($\alpha = \infty$).

both positive and negative externalities associated with players choosing the group-preferred project. Consider a player who considers switching to the group-preferred project. His switching increases the expected coordination payoff for the players with the group-preferred project, as it increases the probability that they interact with players of their own group. On the other hand, the switch lowers the expected coordination payoff to the players with the other project, as there are now fewer players of their group with that project.[14] Since there are more players with the group-preferred project, the positive externality dominates the negative one, and there tends to be too little homophily in equilibrium.

In a stable economic environment, a policy maker may thus want to increase the level of homophily. While the economic incentives for coordination are typically determined by the production technology (and thus outside the reach of the policy maker), we often observe policies to strengthen cultural identity, such as subsidizing cultural programs. Such policies can have both short-term and long-run effects. In the short run, strengthening cultural identity affects the chances that players from the same group coordinate. This effect is given by the partial derivative, $\frac{\partial W^{\infty}}{\partial Q}$, of social welfare with respect to Q. In the long run, the equilibrium level of homophily also adjusts to a new level, and this in turn affects welfare. The long run

[14] A player's choice also affects the payoffs of members of the other group. These effects go in the same direction.

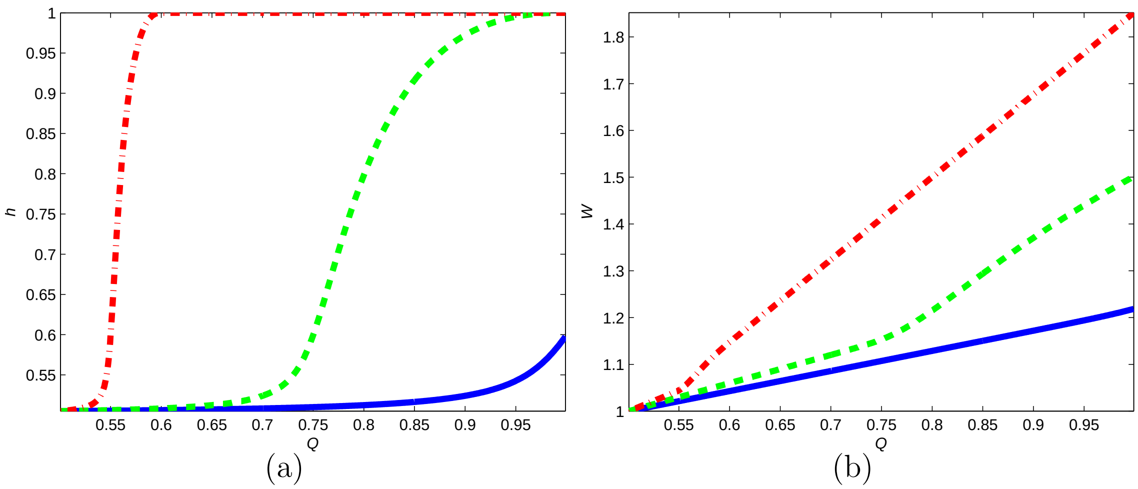

Figure 3: (a) The socially optimal level of homophily as a function of players' cultural identity Q, for $v = 1$ (blue solid line), $v = 2$ (green dashed line), $v = 10$ (red dash-dotted line); (b) Social welfare in equilibrium as a function of Q, again for different values of v. Payoffs are fixed ($\alpha = \infty$) and social welfare levels are normalized such that social welfare is equal to 1 (for all v) when $Q = \frac{1}{2}$.

effect of enhancing cultural identity is thus given by

$$\frac{dW^\infty}{dh} = \frac{\partial W^\infty}{\partial Q} + \frac{\partial W^\infty}{\partial h} \cdot \frac{dh}{dQ}.$$

The next result shows that policies that strengthen cultural identity have positive effects both in the short and the long run.

Corollary 4.5. *Suppose payoffs are fixed (i.e., $\alpha = \infty$). Policies that strengthen cultural identity improve welfare both in the short run and in the long run, and lead to higher levels of homophily.*

The results are illustrated in Figure 3. Strengthening players' cultural identity directly reduces strategic uncertainty by increasing the chance that players from the same group coordinate successfully. In addition, when cultural identity is stronger, the equilibrium level of homophily increases, as players face now aa stronger incentive to interact with their own group (Corollary 4.3). This further reduces strategic uncertainty, as the behavior of members of the own group can be predicted with greater accuracy. Again, this is a robust prediction: even if players' cultural identity cannot be measured, the model still predicts that measures that make the behavior of members of the own group more predictable improves welfare, while leading to more segregation.

These results could explain the popularity of policies that aim to strengthen cultural identity. For example, policies that strengthen an organization's culture are often seen as key to its success (e.g., Tichy and Sherman, 2005). Also, in 19th-century Europe, newly formed nation states built national museums to strengthen national identity (Macdonald, 2012). And social movements in 19th-century U.S. stimulated public school enrollment to build a new, common identity (Meyer et al., 1979). While intuitive, the effects of such policies are difficult to formalize within the standard framework, as the standard framework does not explicitly model players' their cultural identity. In addition, welfare implications may be ambiguous in the standard framework. Since any introspective equilibrium is a correlated equilibrium, there is a correlated equilibrium where strengthening cultural identity improves welfare. But, there are also correlated equilibria in which a majority of the A-players chooses the non-group preferred project b (and analogously for the B-players). When cultural identity is strengthened, the share of A-players choosing project b increases (as the incentive to segregate increases), as in the unique introspective equilibrium. However, unlike in the introspective equilibrium, this may lower social welfare.

4.2. Uncertain economic environments

Oftentimes, the economic environment is not stable, and a policy maker may need to choose a policy before uncertainty is resolved. We consider the case where the potential gains from a technological innovation grows without bounds: as the degree of economic stability α decreases to 1, the expected payoff v^* of coordination on the Pareto superior action goes to infinity. As before, this is a useful benchmark case that allows us to emphasize the driving forces. To gain more insight, it will be instructive to first consider the case where the policy maker cares only about coordination payoffs. In other words, the policy maker does not put any weight on the value that players derive from projects, so that his objective is to maximize $C^\alpha(h)$. Our first result characterizes the level of homophily that maximizes coordination payoffs.

Proposition 4.6. *The level of homophily that maximizes the total coordination payoff $C^\alpha(h)$ converges to*

$$h_C := \frac{1}{8Q-2}$$

as α decreases to 1.

So, the level of homophily that maximizes total coordination payoffs decreases with the strength of players' cultural identity. This is in sharp contrast with the results for stable economic environments. In a stable economic environment, a policy maker that aims to maximize coordination payoffs (as opposed to social welfare) would opt for full segregation, for any value of the parameters.

While the result may appear surprising at first sight, there is a clear intuition. Suppose a player's impulse is to choose the Pareto inferior action. At level 1, he realizes that members of his own group are likely to have a similar impulse. Choosing the Pareto superior action is a best response only if the payoff gain of going against his impulse is high relative to the risk of miscoordinating, that is, if

$$v^* \cdot [\hat{p} \cdot (1 - Q) + (1 - \hat{p}) \cdot \tfrac{1}{2}] \geq v \cdot [\hat{p} \cdot Q + (1 - \hat{p}) \cdot \tfrac{1}{2}]. \tag{4.1}$$

For a player with the non-group preferred project, say, a B-player with project a, the probability that he interacts with a member of his own group is $\hat{p} = 1 - p$ if the level of homophily is $h = p - \frac{1}{2}$. In that case, condition (4.1) becomes:

$$\frac{v^*}{v} \geq \frac{p \cdot \frac{1}{2} + (1 - p) \cdot Q}{p \cdot \frac{1}{2} + (1 - p) \cdot (1 - Q).} \tag{4.2}$$

This holds for a large range of payoff ratios v^*/v if the probability $1 - p$ that the player interacts with his own group is low. Intuitively, at level 1, a player formulates a best response to the belief that other players are following their impulse. Since a player's impulse is not informative of the impulses of members of the other group, the risk of miscoordinating by going against one's impulse is mitigated if there is only a small probability of interacting with the own group. In other words, the pressure to conform is limited. In that case, choosing the Pareto superior action, even if it goes against one's impulse, is a best response already for moderate payoff ratios.

To maximize coordination payoffs, it is not sufficient that a small minority chooses the Pareto superior action; the probability that players with the group-preferred project choose the Pareto superior action should also be substantial. If B-players form a small minority among the players with project a (i.e., $1 - p$ is low), A-players with project a face a high probability of interacting with their own group, so that Eq. (4.1) is hard to satisfy for such players. However, even if these players are unwilling to choose the Pareto superior action at level 1 if it is against their impulse, they may be willing to do so after more introspection. Suppose Eq. (4.2) holds, so that B-players with project a choose the Pareto superior action at level 1. Also suppose that at level 1, it is not a best response for A-players with project a to choose the Pareto superior action if it is against their impulse. At level 2, an A-player with project a thus formulates a best response to the belief that A-players follow their impulse, while B-players (with project a) choose the Pareto superior action regardless of their impulse. Hence, it is a best response for this player to go against his impulse and choose the Pareto superior action if

$$v^* \cdot [p \cdot (1 - Q) + 1 - p] \geq v \cdot [p \cdot Q],$$

where we have used that the probability $\hat{p}$ of interacting with the own group equals p for players with the group-preferred project. This holds if and only if

$$\frac{v^*}{v} \geq \frac{pQ}{1 - pQ}. \tag{4.3}$$

This condition holds for a large range of payoff ratios if the probability $1-p$ that a player with the group-preferred project interacts with a member of the other group is high. Intuitively, while having a small minority (i.e., $1 - p$ low) increases the probability that the minority chooses the Pareto superior action even if it is against their impulse (i.e., Eq. (4.2) is satisfied for a large range of payoff ratios), it decreases the chances that the majority will follow: when the minority is small, members of the majority face a high risk of miscoordinating if they choose the Pareto superior action against their impulse.

A policy maker thus has to balance two factors: the level of homophily has to be such that (1) the minority is willing to choose the Pareto superior action even if it is against their impulse and they expect all other players to follow their impulse (i.e., (4.2) holds); and (2) the majority is willing to follow the minority (i.e., (4.3) holds). The optimal level of homophily in Proposition 4.6 achieves this balance: it ensures that there is a minority that is small enough so that it faces limited pressure to conform, while having the critical mass to influence the majority. This is consistent with work on team composition and creativity. For example, De Dreu and West (2001) and Gibson and Vermeulen (2003) show empirically that minority dissent can make teams more innovative, but only if the minority can influence the decision making process of the majority.

If a policy maker cares not only about coordination payoffs, but also about the value that players derive from projects, the same intuitions apply, but the optimal level of homophily is no longer monotonic in the strength of players' cultural identity:

Proposition 4.7. *In an uncertain economic environment (i.e., $\alpha \downarrow 1$), the socially optimal level of homophily h^α converges to the limiting level h^1. There is $Q^* \in (\frac{1}{2}, 1)$ such that:*

- *If the strength of players' cultural identity is below Q^*, then h^1 is strictly below the level h_C that maximizes coordination payoffs, and increases with Q; and*
- *If Q exceeds Q^*, then the socially optimal level of homophily h^1 and the level of homophily that maximizes coordination payoffs h_C coincide, and decreases with Q.*

So, there are two regimes, as can be seen in Figure 4(a). If cultural identity is strong (i.e., $Q > Q^*$), there is a strong pressure to conform. Any level of homophily that deviates from the

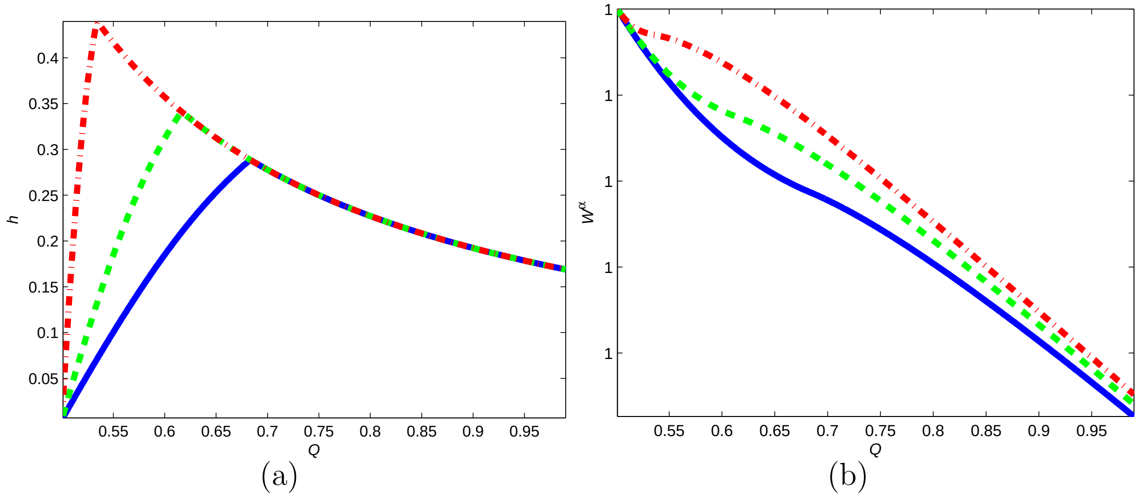

Figure 4: (a) The socially optimal level of homophily as a function of players' cultural identity Q, for $v = 1$ (blue solid line), $v = 2$ (green dashed line), $v = 10$ (red dash-dotted line); (b) Social welfare as a function of Q, again for different values of v. Results are for an uncertain economic environment ($\alpha \downarrow 1$); and social welfare levels are normalized such that social welfare is equal to 1 (for all v) when $Q = \frac{1}{2}$.

level h_C that maximizes coordination payoffs results in a low probability that players choose the Pareto superior action. Hence, it is socially optimal to set the level of homophily equal to the level of homophily that maximizes coordination payoffs. On the other hand, if players' cultural identity is weak (i.e., $Q < Q^*$), then the pressure to conform is not very strong, either for the minority or the majority. This gives room to choose a level of homophily that is lower than h_C, so as to increase the share of players that chooses the project that they intrinsically prefer.

As the coordination payoff v increases, the coordination motive gains in importance relative to the value associated with players choosing the project that they intrinsically prefer. This implies that there is a larger range of parameters for which the level of homophily is chosen such that it maximizes coordination payoffs. This is formalized in the following result.

Corollary 4.8. *The threshold Q^* decreases with the coordination payoff v.*

Figure 4(a) provides an illustration: as the coordination payoff v increases, it is optimal to set the level of homophily equal to the level that maximizes coordination payoffs already when cultural identity is not very strong. This is because the payoffs from coordination now trump the value derived from projects.

Using the characterization in Proposition 4.7, we can ask how social welfare depends on players' cultural identity in an uncertain economic environment.

Corollary 4.9. *In an uncertain economic environment (i.e., $\alpha \downarrow 1$), social welfare decreases with the strength of players' cultural identity.*

The result is illustrated in Figure 4(b). It is again in a marked contrast with the results for stable economic environments. While in a stable economic environment, strengthening cultural identity enhances players' ability to coordinate, in an uncertain economic environment this can be harmful, as it increases the pressure to conform. This means it is harder for a policy maker to achieve a good balance, that is, to choose a level of homophily such that the minority is small enough for it to face limited pressure to conform, yet large enough so that it can influence the majority (i.e., to ensure that Eqs. (4.2) and (4.3) both hold). While the conclusions of Corollaries 4.5 and 4.9 are different, the results really are two sides of the same coin: strengthening players' cultural identity decreases strategic uncertainty, and thus enhances the pressure to conform. In a stable economic environment, this is beneficial, but in an uncertain economic environment, this is not the case.

Proposition 4.7 suggests that it is optimal if the groups are integrated to some degree. Policies that limit contact between individuals, while enhancing contact between others, can be highly successful in shaping social interaction patterns. For example, in schools that limit social choices and have prescribed formats of interaction, the share of intergroup friendships is significantly higher than in schools where students are less restricted in their choice of peers (McFarland et al., 2014). This may motivate policies that assign students to dormitories at random, rather than letting them choose their own roommate (Boisjoly et al., 2006; Burns et al., 2013; Sacerdote and Marmaros, 2006).

Standard equilibrium analysis cannot be used to derive results like the ones presented here. In a standard equilibrium framework, one might argue that the Pareto superior equilibrium is focal, and this is indeed the outcome of a dynamic process that operates through the gradual accretion of precedent (Young, 1993). An advantage of our approach is that it can explain why conformist societies may find it difficult to break out of low-payoff equilibria than more open-minded societies (Mokyr, 1990), and how this interacts with social structure. In a more integrated society or societies with a weak cultural identity, there is more behavioral variation, and this gives the society an opportunity to escape bad equilibria, just like the unintentional errors in the work of Young allow societies to establish efficient conventions.

5. Network formation

In many situations, people can choose how many people they interact with. So, we extend the basic model to allow players to choose how much effort they want to invest in meeting others. We show that the basic mechanisms that drive the tendencies to segregate may be reinforced, and that the model gives rise to network properties that are commonly observed in social and economic networks.

To analyze this setting, it is convenient to work with a finite (but large) set of players.[15] Each group $G = A, B$ has N players, so that the total number of players is $2N$. Players simultaneously choose effort levels and projects in the first stage. They then interact in the coordination game (with fixed payoffs, i.e., $v^* = v$). Effort is costly: a player that invests effort e pays a cost $ce^2/2$. By investing effort, however, a player meets more partners to play the coordination game with (in expectation). Specifically, 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 are matched (and play the coordination game) is

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

where E^π is the total effort of the players with project π.[16] Thus, efforts are complements: players tend to meet each other when they both invest time and resources. This is related to 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.[17] So, the probability of being matched with 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.

As before, at level 0 players choose the project that they intrinsically prefer. So, the probability that a player chooses the group-preferred project 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.[18]

A preliminary result is that the limiting behavior is well-defined, and that it is independent of the choice of effort at level 0.

[15]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.

[16]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.

[17]See, e.g., Cabrales et al. (2011) and Galeotti and Merlino (2014) for applications of this model in economics.

[18]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 2.1).

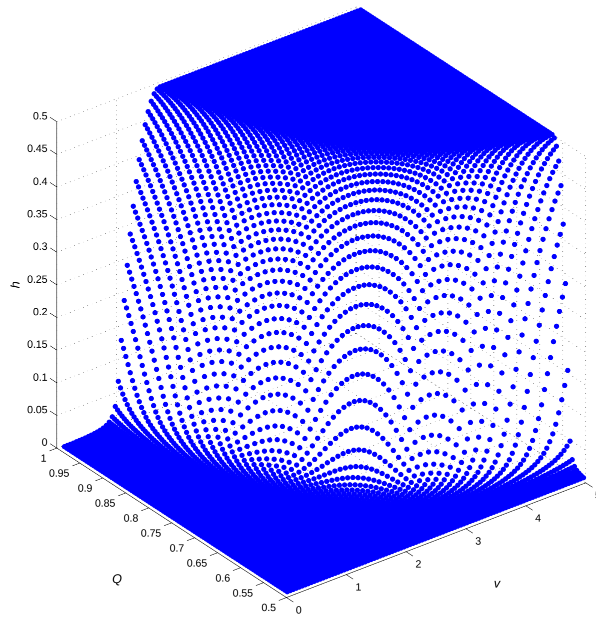

Figure 5: The level of homophily h as a function of the coordination payoff v and the strength of players' cultural identity Q ($c = 1$).

Lemma 5.1. *The limiting probability* p *and the limiting effort choices exist and do not depend on the effort choice at level 0.*

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

Proposition 5.2. *There is a unique introspective equilibrium. In the unique equilibrium, all players choose positive effort. Players that have chosen 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 5 illustrates the comparative statics of the unique equilibrium. As before, the level of homophily increases with the strength of players' cultural identity and with economic incentives, and the two are complements. While the proof of Proposition 5.2 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 5.3. *Consider the limit where the number of players in each group goes to infinity. The effort chosen by the players with the group-preferred project in the unique introspective equilibrium converges to*

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

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

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

which is strictly smaller than the effort e^* (while positive).

Proposition 5.3 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, and this leads to higher levels of 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 meeting someone from the 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 different regimes. If effort costs are small relative to the benefits of coordinating, 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 players from their own group. On the other hand, if effort costs are sufficiently high, then the net benefit of interacting with others is small, even if society is 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:[19]

Proposition 5.4. *Consider the limit where the number of players in each group goes to infinity. 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.*

These results are consistent with empirical evidence. More homogeneous societies have a higher level of social interactions (Alesina and La Ferrara, 2000); and the distribution of the number of connections in social and economic networks has considerable variance (Jackson, 2008). Furthermore, consistent with the theoretical results, friendships are often biased towards own-group friendships, and larger groups form more friendships per capita (Currarini et al., 2009).

Our results put restrictions on the type of networks that can be observed. When relative benefits v/c are high and there is a strong cultural identity Q, networks are dense and are characterized by high levels of homophily and a skewed distribution of the number of connections that players have. Moreover, the network consists of a tightly connected core of players from one group, with a smaller periphery of players from the other group. When v/c increases further, segregation is complete ($h = \frac{1}{2}$), and a densely connected homogenous network results. On the other hand, when economic benefits are limited and cultural identity is weak, networks are disconnected, and feature low levels of homophily and limited variation in the number of connections. Most data on network on social and economic networks is consistent with the case where there is a strong cultural identity and sizeable economic benefits to coordination, with many networks featuring high levels of homophily, a core-periphery structure, high levels of connectedness, and a skewed degree distribution (Jackson, 2008). More research is needed, of course, to establish to what extent these observations can indeed be attributed to the economic and cultural factors that drive players' preferences for reducing strategic uncertainty.

[19]This result follows directly from Proposition 5.2 and Theorem 3.13 of Bollobás et al. (2007). 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 $e^- < e^*$.

6. Conclusions

This paper introduces a novel approach to model players' introspective process, grounded in evidence on Theory of Mind in psychology. We use the framework to show that high levels of homophily are possible when players benefit from reducing strategic uncertainty. This is true even if there are no group-specific externalities and players have no direct preference for interacting with the own group. Modeling players' introspective process explicitly makes it possible to derive unique predictions in a number of different setting and to derive robust and intuitive comparative statics results. Consistent with empirical and experimental evidence, homophily is high when cultural identities are strong, benefits from coordination are large, and networks are formed endogenously. The theory elucidates how the socially optimal level of homophily varies with the economic environment. While segregated societies with a strong cultural identity have an advantage in a stable economic environment, more diverse societies with a weak identity do better in more uncertain environments.

There are a number of directions for future research. On the methodological side, we plan to examine the potential of our approach in general games. In ongoing experimental work, we are investigating to what extent a preference to reduce strategic uncertainty drives homophily. Another promising direction is to study how players' cultural identity coevolves with social structure. Indeed, members of inclusive societies may gain a better understanding of the cultural background of others, while individuals belonging to more segregated societies specialize in their own culture. If that is the case, different social structures may develop depending on initial conditions, and interaction patterns may be persistent, consistent with empirical evidence (Ellison and Powers, 1994). Opening the black box of how intergroup contact affects intergroup understanding may also make it possible to assess which types of interventions are welfare improving: for example, is it important that individuals become sensitive to each other's culture, or is it necessary that they develop a joint identity? To answer these types of questions, it is critical to model cultural identity and players' reasoning processes, and this paper presents a first step in that direction.

Appendix A 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 6. Likewise, on average, players in group B have an intrinsic preference for b.

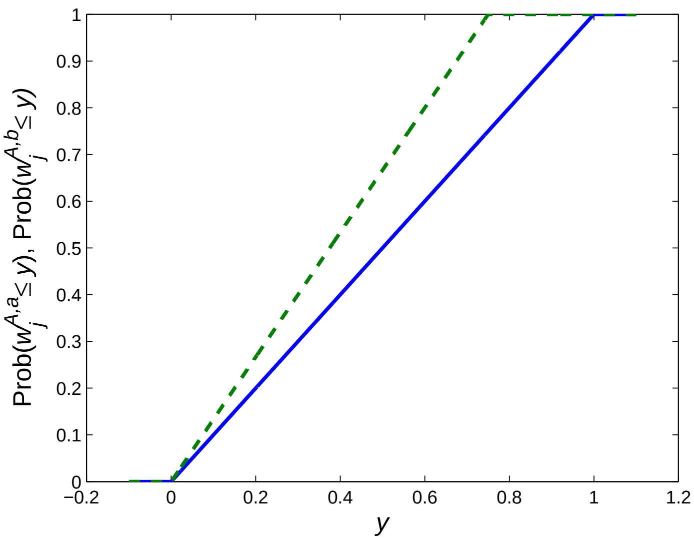

Figure 6: The cumulative distribution functions of $w_i^{A,a}$ (solid line) and $w_i^{A,b}$ (dashed line) for $x = 0.75$.

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 7.

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.

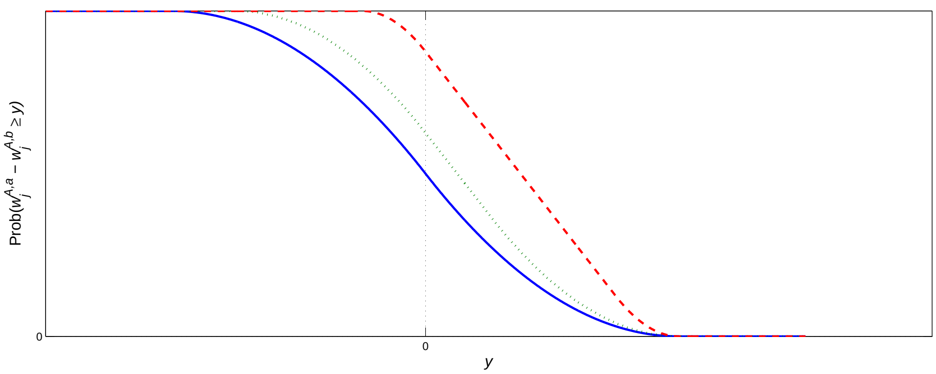

Figure 7: The probability that $w_j^{A,a} - w_j^{A,b}$ is at least y, as a function of y, for $\varepsilon = 0$ (solid line); $\varepsilon = 0.125$ (dotted line); and $\varepsilon = 0.375$ (dashed line).

Appendix B Equilibrium analysis

@@@@Lattice of equilibria? Do we get different comparative statics?

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

$$\left[pQ + (1-p)\cdot\tfrac{1}{2}\right]\cdot v + w_j \geq \left[p\cdot(1-Q) + (1-p)\cdot\tfrac{1}{2}\right]\cdot v + w_j.$$

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

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

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

or, equivalently,

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

where we have defined $\beta := v\cdot(Q - \frac{1}{2})$. 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 A), 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 A), the equilibrium condition reduces to

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

Thus, equilibrium strategies are characterized by a fixed point p of Equation (B.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 (B.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 society can 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.

Appendix C Signaling identity

Thus far, we have assumed that players can choose projects to sort. An alternative way in which people can bias the meeting process is to signal their identity to others. Here, we assume that players can use markers, that is, observable attributes such as tattoos, to signal their identity. This alternative model helps explain why groups are often marked by seemingly arbitrary traits.

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.[20] Each player is matched exactly once.[21] Players' decision to propose or to accept/reject a proposal may depend on project choices, but do 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; if j's proposal was rejected by j', both get a payoff of zero. For simplicity, assume that there are no skill complementarities (i.e., $V = v$).

Players' choices are determined by the introspective process introduced earlier. 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 C.1. *There is a unique introspective equilibrium. In the unique equilibrium, there is complete segregation ($h = \frac{1}{2}$) if and only if*

$$v \cdot (Q - \tfrac{1}{2}) \geq \tfrac{1}{2} - \varepsilon;$$

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

$$\tfrac{1}{2} - \frac{1}{2 - 4\varepsilon}\big(1 - 2\varepsilon - \tfrac{v}{2} \cdot (Q - \tfrac{1}{2})\big)^2.$$

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

Also the comparative statics are similar:

[20] 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.

[21] Such a matching is particularly straightforward to construct when there are finitely many players, as in Section 5. Otherwise, we can use the matching process of Alós-Ferrer (1999) (where the types need to be defined with some care). 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.

Corollary C.2. *The level of homophily h increases with the strength of the cultural identity Q and with the coordination payoff v. Cultural identity and economic incentives are complements: the level of homophily is high whenever the coordination payoff is high and cultural identity is strong.*

So, even if players cannot influence the probability of meeting similar others by locating in a particular neighborhood or joining an exclusive club, they can nevertheless associate preferentially with other members of their own group, provided that they can signal their identity.[22]

Appendix D Complementary skills

Players with different backgrounds may have complementary skills (Page, 2007). To model that players from different groups have complementary skills, we assume that players receive a higher payoff if they successfully coordinate with a member of the other group. That is, payoffs are now given by:

	s^1	s^2
s^1	v,v	0,0
s^2	0,0	v,v

Own group

	s^1	s^2
s^1	V,V	0,0
s^2	0,0	V,V

Other group

where $V > v$. Players follow the same process as before. At level 0, players follow their impulse and select the project they intrinsically prefer. At level $k > 0$, players formulate a best response to actions selected at level $k-1$: a player chooses project a if and only if the expected payoff from a is at least as high as from b, given the choices at level $k-1$.

If the probability that players are matched with an opponent of the same group is $\hat{p} \in (0, 1]$, then a player's expected payoff is

$$\hat{p}Qv + (1 - \hat{p})\tfrac{1}{2}V,$$

so the *marginal benefit of interacting with the own group* is

$$\beta_{CS} := Qv - \tfrac{1}{2}V.$$

Note that β_{CS} can be positive or negative, depending on the relative strengths of players' cultural identity, economic incentives, and skill complementarities. If $\beta_{CS} < 0$, the effect

[22] Unlike 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.

of skill complementarities on payoffs is greater than the benefit of interacting with the own group, and we say that *skill complementarities dominate*. Otherwise, if $\beta_{CS} > 0$, players benefit from interacting with members of their own group as that reduces strategic uncertainty. The analysis from Section 3 extends directly to this case. We therefore focus on the case $\beta_{CS} < 0$ here.

As before, 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. We can prove the analogue of Lemma 3.1 for this setting:[23]

Lemma D.1. *Suppose skill complementarities dominate and that $\beta_{CS} > -\frac{1}{2}$. The limit p^π of the fractions $p_0^\pi, p_1^\pi, \dots$ exists for each project $\pi = a, b$. Moreover, the limits are the same for both projects: $p^a = p^b$.*

Proof. Suppose $\beta_{CS} \in (-\frac{1}{2}, 0)$. By an argument similar to the one in the proof of Lemma 3.1, it follows that the sequence $\{p_k^\pi\}_k$ is weakly decreasing and bounded for every project π. Moreover, $p_k^\pi < \frac{1}{2} + \varepsilon$ for all k. Again, by the monotone convergence theorem, the sequences $\{p_k^a\}_k$ and $\{p_k^b\}_k$ converge to a common limit p. □

The next result shows that there is a unique introspective equilibrium also in this case, and characterizes the equilibrium level of homophily.

Proposition D.2. *Suppose skill complementarities dominate and that $\beta_{CS} > -\frac{1}{2}$. There is a unique introspective equilibrium. The equilibrium fraction of players choosing the group-preferred project is strictly below the initial level (i.e., $h < \varepsilon$), and is given by*

$$h = \frac{\varepsilon}{1 - 2(Qv - \frac{1}{2}V)} > 0.$$

Proof. By Lemma D.1, $p_k \leq p_{k-1}$ for all k. By the monotone sequence convergence theorem, $p = \inf_k p_k$. As before, we can find p by solving the fixed-point equation

$$p = H_\varepsilon(-(2p-1) \cdot \beta_{CS}).$$

Writing $y := -(2p-1) \cdot \beta_{CS}$, we now need to consider two regimes: $y \in (0, \varepsilon)$ and $y \in [2\varepsilon, 1]$ (cf. Appendix A). In the second regime, $H_\varepsilon(y) = \frac{1}{2x}(1-y)^2$, and the fixed-point equation $p(y) = H_\varepsilon(y)$ has two roots y_1, y_2 that lie outside the domain $(0, 2\varepsilon)$. So consider the first

[23] It can be checked that if $\beta_{CS} < -\frac{1}{2}$, then the sequence $p_0^\pi, p_1^\pi, \dots$ does not settle down. Intuitively, players have an incentive to "flee" from players of their own group to reap the high payoffs from interacting with the other group. For example, if A-players make up the majority of players with project a at level k, then even an A-player for whom it is optimal to choose project a at level k may find it beneficial to choose project b at level $k+1$, and A-players form the minority of players with project a at that level.

regime, where $H_\varepsilon(y) = 1 - \frac{1}{2}x - y$. The fixed-point equation has a unique solution y^*, with corresponding limiting probability

$$p = \tfrac{1}{2} - \frac{\varepsilon}{2\beta_{CS} - 1}.$$

It can be checked that p is increasing in β_{CS}, and lies in $(\frac{1}{2}, \frac{1}{2} + \varepsilon)$ for $\beta_{CS} < 0$. $\square$

This is consistent with empirical evidence that skill complementarities across groups can reduce the level of homophily (Aldrich and Kim, 2007). Our model shows how these factors interact: if cultural identities are weak, then complementarities of skills become more important in shaping interactions.

The next result characterizes the socially optimal level of homophily in the presence of skill complementarities.

Proposition D.3. *Suppose skill complementarities dominate. Full segregation is never optimal. The socially optimal level of homophily is given by:*

$$h^* = 4(Qv - \tfrac{1}{2}V)(1 - 2\varepsilon) + 5\varepsilon - 2 + \sqrt{4(Qv - \tfrac{1}{2}V)^2 - 5(Qv - \tfrac{1}{2}V) + 1 + \frac{(Qv - \frac{1}{2}V)}{1 - 2\varepsilon}}.$$

The fraction of players choosing the group-preferred project in the social optimum is below the initial level (i.e., $h^* < \varepsilon$).

This result follows readily from the proof of Proposition 4.2 if we modify the expression $\widetilde{C}^\infty(p)$ for the coordination payoffs to take into account that coordinating with a member of the other group provides payoff $V \geq v$.

Proposition D.3 demonstrates that if skill complementarities dominate, there can be too much homophily in equilibrium. This is consistent with other arguments that show that reducing segregation can improve welfare when there are significant complementarities of skill (e.g. Alesina and La Ferrara, 2005; Ottaviano and Peri, 2006). However, the difference between the socially optimal and equilibrium level of homophily is minimal in our setting, as both are below ε, which can be taken to be arbitrarily small. Intuitively, if skill complementarities dominate, an A-player that chooses the group-preferred project a exerts a negative externality on A-players that choose project a (and a positive one on B-players with project a), as well as a positive externality on A-players that choose project b (and a negative one on B-players that choose b), and likewise for B-players that choose project b (as $\beta_{CS} < 0$). However, in this case, players face a strong incentive to form integrated groups in equilibrium. This means that the share of players experiencing a negative externality is about as large as the share of players experiencing a positive externalities, so that the two types of externalities essentially cancel out.

Appendix E Proofs

E.1 Proof of Proposition 2.1

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

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

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 player follows his impulse. $\square$

E.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

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

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 A).

If the probability that players are matched with an opponent of the same group is $\hat{p} \in (0, 1]$, then a player's expected payoff is

$$v \cdot (\hat{p}Q + (1-\hat{p})\tfrac{1}{2}).$$

The *marginal benefit of interacting with the own group* is thus

$$\beta := v \cdot (Q - \tfrac{1}{2}).$$

As $Q > \frac{1}{2}$, 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 2.1 (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_{k-1}^a, a player from group A chooses project a if and only if

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

This inequality can be rewritten as

$$w_j^{A,a} - w_j^{A,b} \geq -(2p_{k-1}^a - 1) \cdot \beta, \tag{E.1}$$

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

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

where we have used the expression for the tail distribution H_ε from Appendix A. 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 $-(2p_0^\pi - 1) \cdot \beta < 0$. We claim that $p_1^\pi \geq p_0^\pi$ and that $p_1^\pi \in (\frac{1}{2}, 1]$. By the argument above,

$$\begin{aligned}
p_1^\pi &= \mathbb{P}(w_j^{A,a} - w_j^{A,b} \geq -(2p_0^\pi - 1) \cdot \beta) \\
&= H_\varepsilon(-(2p_0^\pi - 1) \cdot \beta) \\
&= \begin{cases} 1 - \frac{1}{2-4\varepsilon} \cdot (1 - 2\varepsilon - (2p_0^\pi - 1) \cdot \beta)^2 & \text{if } (2p_0^\pi - 1) \cdot \beta \leq 1 - 2\varepsilon; \\ 1 & \text{if } (2p_0^\pi - 1) \cdot \beta > 1 - 2\varepsilon; \end{cases}
\end{aligned}$$

where we have used the expression for the tail distribution $H_\varepsilon(y)$ from Appendix A. If $(2p_0^\pi - 1) \cdot \beta > 1 - 2\varepsilon$, the result is immediate, so suppose that $(2p_0^\pi - 1) \cdot \beta \leq 1 - 2\varepsilon$. We need to show that

$$1 - \tfrac{1}{2-4\varepsilon} \cdot (1 - 2\varepsilon - (2p_0^\pi - 1) \cdot \beta)^2 \geq p_0^\pi.$$

Rearranging and using that $p_0^\pi \in (\frac{1}{2}, 1]$, we see that this holds if and only if

$$(2p_0^\pi - 1) \cdot \beta \leq 2 \cdot (1 - 2\varepsilon).$$

But this holds because $(2p_0^\pi - 1)\cdot\beta \le 1-2\varepsilon$ and $1-2\varepsilon \ge 0$. Note that the inequality is strict whenever $\beta < 1-2\varepsilon$, so that $p_1^\pi > p_0^\pi$ in that case.

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

$$p_k^\pi = \begin{cases} 1 - \frac{1}{2-4\varepsilon}\cdot(1-2\varepsilon-(2p_{k-1}^\pi - 1)\cdot\beta)^2 & \text{if } (2p_{k-1}^\pi - 1)\cdot\beta \le 1-2\varepsilon; \\ 1 & \text{if } (2p_{k-1}^\pi - 1)\cdot\beta > 1-2\varepsilon. \end{cases}$$

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

$$1 - \tfrac{1}{2-4\varepsilon}\cdot(1-2\varepsilon-(2p_{k-1}^\pi - 1)\cdot\beta)^2 \ge p_{k-1}^\pi,$$

or, equivalently,

$$2\cdot(1-2\varepsilon)\cdot(1-p_{k-1}^\pi) \ge (1-2\varepsilon-(2p_{k-1}^\pi - 1)\cdot\beta)^2.$$

By the induction hypothesis, $p_{k-1}^\pi \ge p_0^\pi$, so that $1-2\varepsilon \ge 2-2p_{k-1}^\pi$. Using this, we have that $2\cdot(1-2\varepsilon)\cdot(1-p_{k-1}^\pi) \ge 4\cdot(1-p_{k-1}^\pi)^2$. Moreover,

$$(1-2\varepsilon-(2p_{k-1}^\pi - 1)\cdot\beta)^2 \le 4\cdot(1-p_{k-1}^\pi)^2 - 2\beta(1-2\varepsilon)(2p_{k-1}^\pi - 1) + (2p_{k-1}^\pi - 1)^2\beta^2.$$

So, it suffices to show that

$$4\cdot(1-p_{k-1}^\pi)^2 \ge 4\cdot(1-p_{k-1}^\pi)^2 - 2\beta(1-2\varepsilon)(2p_{k-1}^\pi - 1) + (2p_{k-1}^\pi - 1)^2\beta^2.$$

The above inequality holds if and only if

$$(2p_{k-1}^\pi - 1)\beta \le 2\cdot(1-2\varepsilon),$$

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

So, the sequence $\{p_k^\pi\}_k$ 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 $p^a = p^b$. $\square$

E.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 $p_k \ge p_{k-1}$ 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 $\beta \ge 1-2\varepsilon$.

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 A,

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

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

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

Equation (E.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 $\beta > 0$. $\square$

As for the comparative statics in Corollary 4.3, 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 are both positive for any $p < 1$ (and 0 otherwise).

E.4 Proof of Proposition 4.2

Suppose payoffs are fixed, that is, $\alpha = \infty$. We first calculate coordination payoffs and the value derived from projects. It will be convenient to work with payoffs per project. Define $\widetilde{C}^\infty(p)$ to be the total (expected) coordination payoff attained by players with project a when a share p of players choose the group-preferred project. Also, let $\widetilde{\Pi}(p)$ be the total project value for players that choose project a when a share p of players with the strongest intrinsic preference for the group-preferred project choose it. That is, $\widetilde{\Pi}(p)$ is the sum (i.e., integral) of the values $w_j^{A,a}$ of the players j in group A that belong to the share p of the A-players with the strongest intrinsic preference for project a, plus the sum of the values $w_j^{B,a}$ of the players j in group B that belong to the share $1-p$ of the B-players with the strongest intrinsic preference for project a. By symmetry, $\widetilde{C}^\infty(p)$ and $\widetilde{\Pi}(p)$ are equal to the total expected coordination payoffs and the total value derived from projects, respectively, for players with project b. Thus, the total coordination payoff and total value derived from projects (over all projects) as a function of the level of homophily $h = p - \frac{1}{2}$ are equal to $C^\infty(h) = 2\widetilde{C}^\infty(p)$ and $\Pi(h) = 2\widetilde{\Pi}(p)$, respectively, and social welfare is

$$2\Big[\widetilde{C}^\infty(p) + \widetilde{\Pi}(h)\Big]. \tag{E.3}$$

The next two preliminary results characterize $\widetilde{C}^\infty(p)$ and $\widetilde{\Pi}(p)$.

Lemma E.1. *Suppose that payoffs are fixed (i.e., $\alpha = \infty$), and that a share p of players chooses the group-preferred project. The per-project coordination payoff is*

$$\widetilde{C}^\infty(p) := v \cdot \left[Q \cdot \left(p^2 + (1-p)^2\right) + 2 \cdot p \cdot (1-p) \cdot \tfrac{1}{2}\right]. \tag{E.4}$$

Proof. First note that it is never optimal to have less than half the players choose the group-preferred project (i.e., $p < \frac{1}{2}$). To see this, suppose by contradiction that a share $p < \frac{1}{2}$ of players (of a given group, say A) chooses the group-preferred project (say a). Then social welfare increases if the share $1-p$ of players with the strongest preference for the group-preferred project chooses that project (and the other players choose the other project). This does not impact total coordination payoffs (as it does not affect the probability that players interact with a member of their own group), while it increases the share of players that choose the project that they intrinsically prefer.

Let $p \in [\frac{1}{2}, 1]$ be the share of players that have chosen the group-preferred project. Fix a project, say a, and consider an A-player with that project, that is, a player that has chosen the group-preferred project. The expected coordination payoff to such a player is

$$v \cdot [pQ + (1-p)\tfrac{1}{2}];$$

and since the share of A-players with project a is p, the total expected payoff to A-players with project a is

$$p \cdot v \cdot \left[pQ + (1-p)\tfrac{1}{2}\right].$$

Similarly, the expected coordination payoff to a B-player with project a is

$$v \cdot [(1-p)Qv + \tfrac{p}{2}];$$

and the total expected payoff to B-players with project a is

$$(1-p)\cdot v\cdot\left[(1-p)Q+\tfrac{p}{2}\right].$$

Adding all terms together gives $\widetilde{C}^{\infty}(p)$. $\square$

Lemma E.2. *Suppose that players with the strongest preference for the group-preferred project choose that project. When the share of players choosing the group-preferred project equals p, the total value derived from a project is*

$$\widetilde{\Pi}(p):=\begin{cases}\frac{1}{2x}\cdot\left[x+\frac{x^3}{3}-2x\left(1-p-\frac{x}{2}\right)^2+\frac{1}{3}\left(1-p-\frac{x}{2}\right)^3\right] & \textit{if } p\in[\frac{1}{2},\frac{1}{2}+\varepsilon);\\ \frac{1}{2x}\cdot\left[x+\frac{x^3}{3}-x\left(x-\sqrt{2x(1-p)}\right)^2+\frac{2}{3}\left(x-\sqrt{2x(1-p)}\right)^3\right] & \textit{if } p\in[\frac{1}{2}+\varepsilon,1).\end{cases}$$

Proof. To calculate total project value $\widetilde{\Pi}(p)$, fix a group, say A. In the social optimum, all A-players for whom the difference $w_j^{A,a}-w_j^{A,b}$ exceeds a certain threshold y choose project a, and the other A-players choose project b. The share of players for whom $w_j^{A,a}-w_j^{A,b}$ is at least y is given by $p=H_\varepsilon(y)$, where $H_\varepsilon(y)$ is the tail distribution introduced in Appendix A. Since this tail distribution has different regimes, depending on y, we need to consider different cases. Rather than considering different ranges for the threshold y, it will be easier to work with different ranges for $p=H_\varepsilon(y)$.

Case 1: $p\in[\frac{1}{2},\frac{1}{2}+\varepsilon)$. First suppose that the share p of players choosing the group-preferred project lies in the interval $[\frac{1}{2},\frac{1}{2}+\varepsilon)$. As noted above, the threshold $y=y(p)$ solves the equation $p=H_\varepsilon(y)$. It is easy to check that for every $p\in[\frac{1}{2},\frac{1}{2}+\varepsilon)$, the equation $p=H_\varepsilon(y)$ has a solution $y\in[0,2\varepsilon)$, so that (by the definitions in Appendix A) the equation reduces to $p=1-\frac{x}{2}-y$, or, equivalently,

$$y=1-\tfrac{x}{2}-p.$$

For a given $y=y(p)$, if every A-player chooses project a if and only if $w_j^{A,a}-w_j^{A,b}\geq y$, then the share of A-players choosing project a is p. If the A-players with $w_j^{A,a}-w_j^{A,b}\geq y$ choose project a, then their total project value is

$$\tfrac{1}{x}\int_0^x\int_{w_j^{A,a}+y}^1 w_j^{A,a}dw_j^{A,a}dw_j^{A,b},$$

where the factor $1/x$ comes from the uniform distribution of $w_j^{A,b}$ on $[0,x]$. The total project value for A-players that choose project b is given by

$$\tfrac{1}{x}\int_y^x\int_{w_j^{A,a}-y}^x w_j^{A,b}dw_j^{A,b}dw_j^{A,a}+\tfrac{1}{x}\int_0^y\int_0^x w_j^{A,b}dw_j^{A,b}dw_j^{A,a}.$$

The second term is for A-players for whom $w_j^{A,a}$ is so small (relative to the threshold y) that they choose b for any value $w_j^{A,b} \in [0,x]$ (that is, $w_j^{A,a} - y < 0$). The first term describes the total value for A-player for whom $w_j^{A,a} - y \geq 0$. Working out the integrals and summing the terms gives the expression for $\widetilde{\Pi}(p)$ in the lemma for $p \in [\frac{1}{2}, \frac{1}{2} + \varepsilon)$.

Case 2: $p \in [\frac{1}{2} + \varepsilon, 1]$**.** Next suppose $p \in [\frac{1}{2} + \varepsilon, 1]$. Again, fix a group, say A, and note that the A-players for whom $w_j^{A,a} - w_j^{A,b}$ exceeds a threshold $z = z(p)$ choose project a (and the other A-players choose project b). The threshold is again given by the equation $p = H_\varepsilon(z)$, and for $p \in [\frac{1}{2} + \varepsilon, 1]$, this equation reduces to

$$p = 1 - \tfrac{1}{2x}(x + z).$$

It will be convenient to work with a nonnegative threshold, so define $y := -z \geq 0$. Then, rewriting gives[24]

$$y = x - \sqrt{(}2x(1-p)).$$

The total project value for A-players that choose project a (given p) is

$$\tfrac{1}{x}\int_0^y \int_0^1 w_j^{A,a} dw_j^{A,a} dw_j^{A,b} + \tfrac{1}{x}\int_y^x \int_{w_j^{A,b}-y}^1 w_j^{A,a} dw_j^{A,a} dw_j^{A,b},$$

where the first term is for A-players for whom $w_j^{A,b}$ is sufficiently low that they choose project a for any $w_j^{A,a} \in [0,1]$ (given y), and the second term describes the total project value for the other A-players for whom $w_j^{A,a} - w_j^{A,b} \geq -y$, analogously to before. Again, working out the integrals and summing the term gives the expression for $\widetilde{\Pi}(p)$ for $p \in [\frac{1}{2} + \varepsilon, 1]$. $\square$

We are now ready to prove Proposition 4.2. As in the proof of Lemma E.2, we need to consider two cases. We characterize the socially optimal level of homophily both for the case that the marginal benefit of interacting with the own group β is positive, as well as for the case that β is negative. The characterization for this latter case will be useful when we consider complementarities of skills between groups in Appendix D.

Case 1: $p \in [\frac{1}{2}, \frac{1}{2} + \varepsilon)$**.** In this case, the derivative of social welfare with respect to p is given by

$$2 \cdot (2p-1)\beta + \tfrac{1}{2x}\Big[4x(1 - p - \tfrac{x}{2}) - (1 - p - \tfrac{x}{2})^2\Big].$$

[24]Note that $y' = x + \sqrt{(}2x(1-p))$ also solves the equation. However, a threshold $z' = -y$ less than $-x$ is not feasible: it corresponds to a share of players that choose the group-preferred project that is greater than 1.

Setting the derivative equal to 0 and solving for p gives two roots:

$$r_1 = 4\beta x - \tfrac{5x}{2} + 1 + \sqrt{4\beta^2 - 5\beta + 1 + \tfrac{\beta}{x}},$$

and

$$r_2 = 4\beta x - \tfrac{5x}{2} + 1 - \sqrt{4\beta^2 - 5\beta + 1 + \tfrac{\beta}{x}}.$$

It is straightforward to verify that $r_2 \le \frac{1}{2}$ whenever $x \ge \frac{1}{9}$. Also, if $x \ge \frac{1}{9}$, the root r_1 lies in $[\frac{1}{2}, \frac{1}{2} + \varepsilon)$ if and only if $\beta < 0$. It can be checked that the second-order conditions are satisfied, so $h^* = r_1 - \frac{1}{2}$ is the optimal level of homophily if $\beta < 0$.

Case 2: $p \in [\frac{1}{2} + \varepsilon, 1]$. In this case, the derivative is

$$2 \cdot (2p-1)\beta + \sqrt{2x(1-p)} - x.$$

Again, the first-order condition gives two solutions:

$$r_1' = \tfrac{1}{2} + \frac{x}{4\beta^2}\left[\beta - \tfrac{1}{4} + \sqrt{\tfrac{\beta^2}{x} - \tfrac{\beta}{2} + \tfrac{1}{16}}\right],$$

and

$$r_2' = \tfrac{1}{2} + \frac{x}{4\beta^2}\left[\beta - \tfrac{1}{4} - \sqrt{\tfrac{\beta^2}{x} - \tfrac{\beta}{2} + \tfrac{1}{16}}\right].$$

For any combination of parameters, $r_2' \le \frac{1}{2}$. Clearly, $r_1' > r_2'$; moreover, r_1' is a saddle point (and thus a point of inflection) if and only if $2\beta \ge x$. If $2\beta \ge x$, then the derivative of social welfare with respect to p is positive in the neighborhood of r_1'. In that case, social welfare attains its maximum at the boundary $p = 1$, and the optimal level of homophily is $h^* = 1 - \frac{1}{2} = \frac{1}{2}$. If $2\beta \in (0, x)$, then $r_1' \in (\frac{1}{2} + \varepsilon, 1]$, and conversely, if $r_1' \in [\frac{1}{2} + \varepsilon, 1]$, then $\beta \in (0, \frac{x}{1-x}]$. Hence, if $2\beta \in (0, x)$, the optimal level of homophily is $h^* = r_1' - \frac{1}{2} > \varepsilon$. □

E.5 Proof of Lemma 5.1

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 A) 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[25]

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

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

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

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

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

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

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

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).

[25] 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.

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

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

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 A), 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

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

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 $N_{k-1}^{A,a}$ of A-players with project b at level $k-1$ and the number $N_{k-1}^{B,a}$ 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_m^D;
- 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 $e_{k-1}^D > e_{k-1}^M > 0$ for $k \geq 2$;
- we have $e_{k-1}^D - e_{k-1}^M \geq e_{k-2}^D - e_{k-2}^M$.

We write N_{k-1}^D and N_{k-1}^M 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

$$e_k^D := \left(\frac{v}{c}\right) \cdot \mathbb{E}\left[\frac{N_{k-1}^D \cdot e_{k-1}^D \cdot Q + N_{k-1}^M \cdot e_{k-1}^M \cdot \frac{1}{2}}{N_{k-1}^D \cdot e_{k-1}^D + N_{k-1}^M \cdot e_{k-1}^M}\right],$$

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

$$e_k^M := \left(\frac{v}{c}\right) \cdot \mathbb{E}\left[\frac{N_{k-1}^M \cdot e_{k-1}^M \cdot Q + N_{k-1}^D \cdot e_{k-1}^D \cdot \frac{1}{2}}{N_{k-1}^D \cdot e_{k-1}^D + N_{k-1}^M \cdot e_{k-1}^M}\right].$$

Again, it is easy to verify that

$$e_k^D + e_k^M = \frac{v}{c}(Q + \tfrac{1}{2}). \tag{E.5}$$

We claim that $e_k^D \geq e_{k-1}^D$ (so that $e_k^M \leq e_{k-1}^M$). It then follows from the induction hypothesis that $e_k^D > e_k^M$ and that $e_k^D - e_k^M \geq e_{k-1}^D - e_{k-1}^M$.

To show this, recall that for $m = 1, \ldots, k-1$, we have that $e_m^D > e_m^M$ and $e_m^D + e_m^M = \frac{v}{c}(Q + \frac{1}{2})$. Define

$$g_{k-1}^D(N_{k-2}^D, N_{k-2}^M, e_{k-2}^D) \quad := \quad \left(\frac{v}{c}\right) \cdot \left(\frac{N_{k-2}^D \cdot e_{k-2}^D \cdot Q + N_{k-2}^M \cdot e_{k-2}^M \cdot \frac{1}{2}}{N_{k-2}^D \cdot e_{k-2}^D + N_{k-2}^M \cdot e_{k-2}^M}\right)$$
$$g_k^D(N_{k-1}^D, N_{k-1}^M, e_{k-1}^D) \quad := \quad \left(\frac{v}{c}\right) \cdot \left(\frac{N_{k-1}^D \cdot e_{k-1}^D \cdot Q + N_{k-1}^M \cdot e_{k-1}^M \cdot \frac{1}{2}}{N_{k-1}^D \cdot e_{k-1}^D + N_{k-1}^M \cdot e_{k-1}^M}\right)$$

so that e_{k-1}^D and e_k^D are just proportional to the expectation of g_{k-1}^D and g_k^D (over N_{k-1}^D and N_{k-1}^M), respectively, analogous to before. It is easy to verify that $g_k^D(N_{k-1}^D, N_{k-1}^M, e_{k-1}^D) \geq g_k^D(N_{k-1}^D, N_{k-1}^M, e_{k-1}^M)$. Consequently,

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

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

$$e_k^D \geq \left(\frac{v}{c}\right) \cdot \mathbb{E}\left[\frac{N_{k-1}^D \cdot Q + N_{k-2}^M \cdot \frac{1}{2}}{N_{k-1}^D + N_{k-2}^M}\right]. \tag{E.6}$$

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

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

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

$$e_{k-1}^D \leq \left(\frac{v}{c}\right) \cdot \mathbb{E}\left[\frac{N_{k-1}^D \cdot Q + N_{k-1}^M \cdot \frac{1}{2}}{N_{k-1}^D + N_{k-1}^M}\right]. \tag{E.7}$$

The result now follows by comparing Equations (E.6) and (E.7). Also, using that g_k^D is increasing and decreasing in its first and second argument, respectively, we have that

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

and it follows from (E.5) that $e_k^D, e_k^M \in [\frac{v}{2c}, \frac{v \cdot Q}{c}]$.

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

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

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 (E.5) 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 $e^M := \lim_{k\to\infty} e_k^M = \frac{v}{c}(Q + \frac{1}{2}) - e^D$). □

E.6 Proof of Proposition 5.2

Recall the definitions from the proof of Lemma 5.1. 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

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

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 5.1, we have $0 < e^M < e^D < \frac{v}{c}(Q + \frac{1}{2})$. Moreover, $e^D + e^M = \frac{v}{c}(Q + \frac{1}{2})$.

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

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

so that $e^D = h^D(e^D)$ in the introspective equilibrium.[26] Since $e^D + e^M = \frac{v}{c}(Q + \frac{1}{2})$ and $e^M > 0$, we have $e^D \in (0, \frac{v}{c}(Q + \frac{1}{2}))$. It is easy to check that $\lim_{e^D \downarrow 0} h^D(e^D) = \frac{v}{2c} > 0$ and

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

that $\lim_{e^D \uparrow \frac{v}{c}(Q+\frac{1}{2})} h^D(e^D) = \frac{vQ}{c} < \frac{v}{c}(Q + \frac{1}{2})$. 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

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

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 $h^D(e^D)$:[27]

$$\frac{d^2 h^D(e^D)}{de^D} = \frac{2v^2}{c^2} \cdot \left(Q^2 - \frac{1}{4}\right) \sum_{n^D=1}^{N} \binom{N}{n^D} p^{n_D}(1-p)^{N-n^D}$$
$$\sum_{n^M=1}^{N} \binom{N}{n^M} p^{N-n_M}(1-p)^{n^M} \cdot \frac{n^D n^M (n^M - n^D)}{(n^D \cdot e^D + n^M \cdot e^M)^3}.$$

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

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

We can rewrite this condition as

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

But this is equivalent to the inequality

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

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

[27] 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 $n^D > 0, n^M = 0$.

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 5.1). 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. □

E.7 Proof of Proposition 5.3

Recall the notation introduced in the proof of Lemma 5.1. 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 5.2) converges (pointwise) to

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

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 5.2. 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 5.2), and where we have used that $h = p - \frac{1}{2}$). □

References

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

Aldrich, H. E. and P. H. Kim (2007). Small worlds, infinite possibilities? How social networks affect entrepreneurial team formation and search. *Strategic Entrepreneurship Journal 1*, 147–165.

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

Alesina, A. and E. La Ferrara (2005). Ethnic diversity and economic performance. *Journal of Economic Literature 43*, 762–800.

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

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

Apperly, I. (2012). *Mindreaders: The Cognitive Basis of "Theory of Mind"*. Psychology Press.

Ashford, B. E. and S. A. Johnson (2014). Which hat to wear? The relative salience of multiple identities in organizations. In M. A. Hogg and D. J. Terry (Eds.), *Social Identity Processes in Organizational Contexts*, Chapter 3. Psychology Press.

Ashford, B. E. and D. Mael (1989). Social identity theory and the organization. *Academy of Management Review 14*, 20–39.

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

Bacharach, M. and D. O. Stahl (2000). Variable-frame level-n theory. *Games and Economic Behavior 32*, 220–246.

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

Baron-Cohen, S., M. Lombardo, H. Tager-Flusberg, and D. Cohen (Eds.) (2013). *Understanding Other Minds: Perspectives from Developmental Social Neuroscience.* Oxford University Press.

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.

Bénabou, R. (1996). Heterogeneity, stratification, and growth: Macroeconomic implications of community structure and school finance. *American Economic Review 86*, 584–609.

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

Boisjoly, J., G. J. Duncan, M. Kremer, D. M. Levy, and J. Eccles (2006). Empathy or antipathy? The impact of diversity. *American Economic Review 96*, 1890–1905.

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.

Burns, J., L. Corno, and E. La Ferrara (2013). Does interaction affect racial prejudice and cooperation? Evidence from randomly assigned peers in South Africa. Working paper.

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

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.

Cornell, B. and I. Welch (1996). Culture, information, and screening discrimination. *Journal of Political Economy 104*, 542–571.

Crawford, V. P. (2007). Let's talk it over: Coordination via preplay communication with level-k thinking. Working paper, U.C. San Diego.

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.

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 (2013). Identity, homophily and in-group bias. Working paper, University of Essex.

Cutler, D. and E. Glaeser (1997). Are ghettos good or bad? *Quarterly Journal of Economics CXII*, 827–872.

De Dreu, C. K. W. and M. A. West (2001). Minority dissent and team innovation: The importance of participation in decision making. *Journal of Applied Psychology 86*, 1191–1201.

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

Durlauf, S. N. (1996). A theory of persistent income inequality. *Journal of Economic Growth 1*, 75–93.

Ellingsen, T. and R. Östling (2010). When does communication improve coordination? *American Economic Review 100*, 1695–1724.

Ellison, C. C. and D. A. Powers (1994). The contact hypothesis and racial attitudes among Black Americans. *Social Science Quarterly 75*, 385–400.

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.

Fiske, S. T. and S. E. Taylor (2013). *Social Cognition: From Brains to Culture*. SAGE.

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

Gibson, C. and F. Vermeulen (2003). A healthy divide: Subgroups as a stimulus for team learning behavior. *Administrative Science Quarterly 48*, 202–239.

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. (1993). Contract enforceability and economic institutions in early trade: The Maghribi traders' coalition. *American Economic Review 83*, 525–548.

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.

Hogg, M. A. (2007). Uncertainty-identity theory. Volume 39 of *Advances in Experimental Social Psychology*, pp. 69 – 126. Academic 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.

Jacquemet, N. and C. Yannelis (2012). Indiscriminate discrimination: A correspondence test for ethnic homophily in the Chicago labor market. *Labour Economics 19*, 824–832.

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

Jenkins, R. (2014). *Social Identity* (Fourth ed.). Routledge.

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

Kant, I. (1781/1997). *The critique of pure reason.* Cambridge University Press.

Kimbrough, E. O., N. Robalino, and A. J. Robson (2013). The evolution of "Theory of Mind": Theory and experiments. Working paper 1907, Cowles Foundation.

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.

Lazear, E. P. (1999). Culture and language. *Journal of Political Economy 107*(S6), S95–S126.

Locke, J. (1690/1975). *An essay concerning human understanding.* Oxford University Press.

Macdonald, S. J. (2012). Museums, national, postnational and transcultural identities. In B. M. Carbonell (Ed.), *Museum Studies: An Anthology of Contexts*, Chapter 29. John Wiley & Son.

Marsden, P. V. (1988). Homogeneity in confiding relations. *Social Networks 10*, 57–76.

McFarland, D. A., J. Moody, D. Diehl, J. A. Smith, and R. J. Thomas (2014). Network ecology and adolescent social structure. *American Sociological Review*. Forthcoming.

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.

Meyer, J. W., D. Tyack, J. Nagel, and A. Gordon (1979). Public education as nation-building in America: Enrollments and bureaucratization in the American states, 1870–1930. *American Journal of Sociology 85*, 591–613.

Mill, J. S. (1872/1974). *A system of logic, ratiocinative and inductive*, Volume 7 of *Collected works of John Stuart Mill*. University of Toronto Press.

Mokyr, J. (1990). *The Lever of Riches: Technological Creativity and Economic Progress*. Oxford University Press.

Morris, S. and H. S. Shin (2003). Global games: Theory and applications. In M. Dewatripont, L. P. Hansen, and S. J. Turnovsky (Eds.), *Advances in economics and econometrics: Eighth World Congress*, Volume 1. Cambridge University Press.

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

Nagel, J. (1994). Constructing ethnicity: Creating and recreating ethnic identity and culture. *Social Problems 41*, 152–176.

Ottaviano, G. I. P. and G. Peri (2006). The economic value of cultural diversity: Evidence from US cities. *Journal of Economic Geography 6*, 9–44.

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

Park, J. J., N. Denson, and N. A. Bowman (2013). Does socioeconomic diversity make a difference? Examining the effects of racial and socioeconomic diversity on the campus climate for diversity. *American Educational Research Journal 50*, 466–496.

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

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

Phelps, E. S. (1972). The statistical theory of racism and sexism. *American Economic Review 62*, 659–661.

Russell, B. (1948). *Human knowledge: Its scope and limits*. George Allen & Unwin.

Sacerdote, B. and D. Marmaros (2006). How do friendships form? *Quarterly Journal of Economics 121*, 79–119.

Schelling, T. (1971). Dynamic models of segregation. *Journal of Mathematical Sociology 1*, 143–186.

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

Sethi, R. and M. Yildiz (2014). Perspectives, opinions, and information flows. Working paper, Barnard College and MIT.

Skyrms, B. (1990). *The Dynamics of Rational Deliberation.* Harvard University Press.

Staber, U. (2001). The structure of networks in industrial districts. *International Journal of Urban and Regional Research 25*, 537–552.

Tichy, N. M. and S. Sherman (2005). *Control Your Destiny or Someone Else Will: How Jack Welch is Making General Electric the World's Most Competitive Company.* Harper Collins.

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.

Young, H. P. (1993). The evolution of conventions. *Econometrica 61*, 57–84.