

# A belief-based theory of homophily<sup>\*</sup>

Willemien Kets<sup>†</sup>

Alvaro Sandroni<sup>‡</sup>

February 15, 2015

## 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. Whether homophily is socially optimal depends crucially on the degree of economic stability.

---

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

<sup>†</sup>Kellogg School of Management, Northwestern University. E-mail: w-kets@kellogg.northwestern.edu

<sup>‡</sup>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, 2014, 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). We provide a theory of homophily that does not assume homophilous preferences. Rather, these preferences are a natural outcome of individuals' desire to reduce uncertainty in strategic settings. Our framework makes it possible to evaluate different policy interventions and to derive clear and intuitive comparative statics.

Our starting point is that to understand homophily we need to unlook the black box of cultural identity. Following Kreps (1990), we view culture as a means to reduce strategic uncertainty. To fix ideas, consider a manager and an employee. The manager can choose a direct or indirect communication style, and the employee can choose a mode of interpretation. While direct communication is appropriate in some settings, sometimes tact is called for. When there is uncertainty about the appropriate behavior, cultural rules can act as focal principles. If the manager and employee have a similar background, they will typically agree on what mode of communication is appropriate in a given context, and they will be able to communicate effectively.

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 for action. However, there may be a focal point in this game, which may depend on the context of the game. If players share a cultural code, they may successfully coordinate by inferring the same focal point from the context of the game.

To formally model these ideas and connect them to the problem of homophily, we build on the dual process account of Theory of Mind in psychology.<sup>1</sup> The dual process account of Theory of Mind posits that an individual start with instinctive reactions and then adapt his views by reasoning about what he would do if he were in the opponent’s position. 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. By introspection, the player realizes that he may act on instinct and so, his opponent may also act on instinct. In addition, he reasons that if his opponent is similar to him that the impulse of the opponent may also 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*.

In line with neurological and behavioral evidence (de Vignemont and Singer, 2006; Jackson and Xing, 2014),<sup>2</sup> players find it easier to put themselves into the shoes of players that are similar to them. Formally, players belong to different groups, and initial impulses are (imperfectly) correlated within groups and independent across groups. So, players from the same group are more likely to agree on which action is focal.

Our first result shows that there is an unique introspective equilibrium, and in this equilibrium each player follows his initial impulse. So, the naive response of following one’s initial impulse is, in fact, 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 similar players coordinate more effectively. This provides an incentive to seek out similar people, that is, to be *homophilous*.<sup>3</sup>

One way players can enhance the chance of interacting with each other is to choose the same project (e.g., a hobby, profession, or neighborhood). So, we consider 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 finally

---

<sup>1</sup>See Epley and Waytz (2010) for a survey. The dual process account of Theory of Mind relies on a rapid instinctive process and a slower cognitive process. So, it is related to the two-systems account of decision-making under uncertainty, popularized by Kahneman (2011). The foundations of dual processes theory go back to the work of the psychologist William James (1890/1983). See Section 2 for an extended discussion.

<sup>2</sup>Empirical and experimental evidence also shows that people trust people that are similar to them more (DeBruine, 2002; Guiso et al., 2009).

<sup>3</sup>Alternatively, players could reduce the risk of miscoordinating by learning the cultural code of the other group (Lazear, 1999). However, this may be costly.

modify them through higher-order reasoning. We show that there is a unique introspective equilibrium. In the unique equilibrium, players from the same group overwhelmingly choose the same project, regardless of their intrinsic sentiments over projects. 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). Thus, introspection and reasoning lead to effective coordination as well as homophily.

Our results show that the root cause of homophily may not be a direct taste for interacting with similar others, but rather reflects a preference to reduce strategic uncertainty. This approach is consistent with the observation that segregation is often based on traits that are behaviorally relevant (e.g., profession, religion) rather than traits that are less tied to behavior, such as height (McPherson et al., 2001). This allows us to shed light on various empirical regularities, as discussed in Section 3. Our results may also explain why homophily is observed in some settings, but not in others, even if the underlying game is similar. For example, some organizational cultures do not strongly guide initial impulses, which leads to less segregated networks (Staber, 2001). Consistent with these observations, our model predicts that the level of homophily is lower when cultural identity is weaker. Finally, regardless of the distribution over impulses, when coordination payoffs are high, the level of homophily is necessarily high. This is a testable implication of our model that does easily follow from standard game-theoretical models.

We now turn to the question of optimal social structure. We allow for uncertainty about the underlying game. Either the game is the pure coordination game considered above, or one of the equilibria Pareto dominates the other. We distinguish two cases. The first case concerns a stable economic environment: with high probability, the game is close to the pure coordination game. The second case considers an uncertain economic environment where there is a significant chance that an innovation significantly enhances the payoffs in the Pareto-dominant equilibrium. We investigate how the social structure that maximizes social welfare depends on the economic environment.

If the economic environment is stable, it is socially optimal to have high levels of homophily. Moreover, strengthening players' cultural identity reduces strategic uncertainty and increases social welfare, even though it leads to segregation. This may be the purpose of policies that strengthen cultural identity 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. In this case, social welfare is maximal if there is a minority of a significant size. 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 conform to the (expected) action choice by members of his group. 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. As before, strengthening players' cultural identity reduces strategic uncertainty, but the welfare effects are now reversed: social welfare is lower if cultural identity is stronger. This is because 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 initial impulse. This 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.

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 coordination also limits its flexibility when potentially Pareto improving innovations are likely. By contrast, more integrated societies with a weaker cultural identity have the agility to perform well in such an uncertain environment, consistent with historical evidence (Mokyr, 1990). As we discuss in Section 4, these results cannot be obtained using the standard framework. In particular, the standard framework cannot explain how the optimal social structure depends on the economic environment.

Formally modeling a mechanism that can capture these common intuitions has several benefits. First, 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. Moreover, it allows us to characterize the conditions under which homophily is socially optimal.

Thus far, we have assumed that all players are matched with exactly one other player. Section 5 captures 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). Other extensions of the model consider the effects of skill complementarities between groups (Appendix D) and the ability to use markers, such as

tatoos or business attire, as a way to signal cultural identity (Appendix C).

The heart of our contribution lies in the modeling of players' reasoning process. In a setting where standard equilibrium refinements have no bite, an explicit model of players' reasoning process is a powerful method to obtain uniqueness in a range of different settings.<sup>4</sup> This equilibrium uniqueness permits intuitive yet subtle comparative statics, and allows for direct welfare analysis 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.<sup>5</sup> 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), so that it is not the case that all players have a preference to interact with the members of the same group (e.g., the high-skilled group). Again, this greatly expands the range of settings where homophily and segregation can

---

<sup>4</sup>The multiple equilibria in games typically used in models homophily may have very different properties (Appendix B). Multiplicity of equilibrium is sometimes dealt with by focusing on equilibria that satisfy stability properties (e.g., Alesina and La Ferrara, 2000; Bénabou, 1993; Sethi and Somanathan, 2004), but such refinements do not always produce uniqueness.

<sup>5</sup>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.

be prominent phenomena, and allows us to analyze measures aimed at influencing interactions when no group is universally seen as a 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 non-equilibrium behavior. Indeed, we show that modeling players' reasoning process can give rise to new insights even if one focuses on the equilibrium limit. In addition, 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.<sup>6</sup>

Our work sheds light on experimental findings that social norms and group identity can lead to successful 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  $\hat{p}$ . Matched players 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 for the group (in

---

<sup>6</sup>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.

the given context). So, 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 analogous 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 minor impact 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 introspective process continues to higher orders: at level  $k > 1$ , players formulate a best response to their beliefs about their opponents' action at level  $k - 1$ . Together, this constitutes a reasoning process of increasing levels. These levels do not represent actual behavior; they are merely constructs in a player's mind. We are interested in the limit of this process as the level  $k$  goes to infinity. If such a limit exists for each player, then the profile of such limiting strategies is referred to as an *introspective equilibrium*.

Our approach is motivated by the dual process account of 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. This reasoning process 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).<sup>7,8</sup> A critical assumption is that players' impulses are correlated (perhaps slightly) within groups (i.e.,  $q > \frac{1}{2}$ ), that is, a player's own impulse is informative of the impulses of players that are similar to him. 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 results 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 her impulse, it is optimal for him to do so as well; this, in turn, makes it optimal for the opponent to follow her impulse.

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.<sup>9</sup> By contrast, the introspective process selects a unique equilibrium. This uniqueness will prove critical for the comparative static results in the next sections.

While it is natural to assume that players follow their impulse at level 0, our results do not depend on this. As long as each player is more likely than not to follow his impulse, our result continues to hold. A core assumption is that players do not have a strong predisposition to choose a fixed action, regardless of context. Also, the result does not hinge on the states of

---

<sup>7</sup>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."

<sup>8</sup>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.

<sup>9</sup>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.

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 actions).

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, 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:

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

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.*

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,<sup>10</sup> 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

---

<sup>10</sup>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.

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 each  $A$ -player  $j$ , the value  $w_j^{A,a}$  of project  $a$  is drawn uniformly at random from  $[0, 1]$ , while the value  $w_j^{A,b}$  of project  $b$  is drawn uniformly at random from  $[0, 1 - 2\varepsilon]$ , for 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. As players follow their impulse when playing the coordination game (Proposition 2.1), the probability that a player successfully coordinates with a member of her own group is  $Q$ , while the probability that she coordinates successfully with a member of the other group is  $\frac{1}{2}$ . So, if the probability of interacting with a member of the own group is  $\hat{p}$ , then the expected payoff of an  $A$ -player with project  $a$  is

$$v \cdot [\hat{p} \cdot Q + (1 - \hat{p}) \cdot \frac{1}{2}] + w_j^{A,a},$$

and likewise for other projects and groups.

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  $p^\pi$  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:  $p^a = p^b$ .*

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, players end up following their impulses after they have gone through the entire reasoning process. In contrast, in the case of endogenous matching, players may not act on impulse. Intuitively, at level 1, players 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. For some parameters, all players choose the group-preferred project. This includes those who have a strong intrinsic preference for the alternative project, and, hence, would instinctively choose the alternative 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.,  $\varepsilon$  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*

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

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

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

*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, group preferences over projects are arbitrarily close (i.e.,  $\varepsilon$  small). It follows that people that share a common background have incentives to become similar on other dimensions as well, e.g., by choosing the same hobbies, professions, or clubs as other members of their group, consistent with empirical evidence (Kossinets and Watts, 2009). Note that 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.

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: homophily is high when either cultural identity or the coordination payoff is high.*

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 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 some 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 standard equilibrium analysis. In contrast, in this model, the majority 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 can also help understand the strong use of distinguishing characteristics, even those that are negatively valued (Ashford and Mael, 1989). Also, the desire to reduce strategic uncertainty may help explain the interaction patterns that are observed when identities are nested (Nagel, 1994; Ashford and Johnson, 2014). Individuals may seek those that match a narrow identity in some cases (e.g., Korean-Americans vs. Asian-Americans); and a broader one in other cases (e.g., Asian-Americans vs. Americans at large). 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 does not change. 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 is typically not the result of 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 the observed phenomena. More fundamentally, it would not allow one to predict interaction patterns from primitives *ex ante*.

Our results do not depend on our specific assumptions, such as the exact assumptions on preferences or the signal structure. 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. 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. Moreover, similar results obtain when players cannot sort by choosing projects, but instead can signal their identity. As we show in Appendix C, we show that our results go through if players choose markers, that is, observable attributes such as tattoos or specific attire, to signal their identity and enhance their chances to meet similar others. 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 have complementary skills. Our model can easily accommodate this possibility, by assuming that 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 skill complementarities, conditional on successful coordination. In Appendix D, we characterize this tradeoff, and show that there is significant homophily in equilibrium if and only if the gains from skill complementarities are limited. Since the effects of skill complementarities are entirely expected, we abstract from it in the remainder of the paper.

## 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 how the optimal level of homophily changes with the economic environment. Consider a policy maker who can allocate players to projects and thus chooses the level of homophily.<sup>11</sup> The policy maker aims to maximize social welfare (i.e., the sum of coordination payoffs and project values), but faces uncertainty about payoffs at the time he chooses a policy. Specifically, 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. The policy maker chooses an ex ante optimal policy, that is, he chooses a policy that maximizes social welfare before payoffs are realized. After payoffs are realized, players observe the payoffs  $v$  and  $v^*$  and play the coordination game, taking project assignments (and the resulting 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.<sup>12</sup>

We contrast the case where it is likely that  $v^*$  and  $v$  are close and the case where there is a significant chance that  $v^*$  is much higher than  $v$ . We capture this by assuming that the ratio  $v^*/v$  of 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  $\alpha$  is large, the payoff  $v^*$  equals  $v$  with

---

<sup>11</sup>Policy measures that affect the level of homophily include instituting inclusionary zoning practices, building housing projects for mixed-income communities, or allocating students to dorm rooms or classes. One can also think of the policy maker as a manager who decides whom to hire for his team.

<sup>12</sup>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 a random time.

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 and are identical to each other:  $v^* = v$  with probability 1. On the other hand, if  $\alpha$  is close to 1, then the expected value of  $v^*$  grows arbitrarily large, and we say that the economic environment is uncertain. Thus, the parameter  $\alpha$  measures the degree of economic stability.

Clearly, for a given level of homophily, it is socially optimal to assign the players with the highest intrinsic preferences for a project to that project. Thus, if the share of players with the group-preferred project is  $p = h + \frac{1}{2}$  in a social optimum, then the share  $p$  of players with the highest intrinsic preferences for the group-preferred project are assigned to that project. That means that we can characterize the social optimum by the level of homophily  $h$ . The *socially optimal level of homophily*  $h^\alpha$  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.

#### 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.** *As  $\alpha$  goes to infinity, the socially optimal level of homophily converges to the socially optimal level of homophily when payoffs are fixed, that is,  $h^\alpha$  converges to  $h^\infty$  as  $\alpha \rightarrow \infty$ . Moreover, social welfare also converges:  $W^\alpha(h^\alpha) \rightarrow W^\infty(h^\infty)$ .*

So, we can concentrate on the benchmark case of fixed payoffs, that is,  $v^* = v$ . By Proposition 4.1, the results hold approximately for stable economic environments. The next result fully 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*

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

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

$$h^\infty = \frac{(1 - 2\varepsilon)}{4v^2(Q - \frac{1}{2})^2} \cdot \left[ v(Q - \frac{1}{2}) - \frac{1}{4} + \sqrt{\frac{4v^2(Q - \frac{1}{2})^2}{1 - 2\varepsilon} - \frac{v}{2}(Q - \frac{1}{2}) + \frac{1}{16}} \right].$$

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  $h^\infty$  increases with the strength of the cultural identity  $Q$  and with the coordination payoff  $v$ ; as with the equilibrium level of homophily, cultural identity and economic incentives are complements in determining the socially optimal level of homophily.*

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.

We can now compare the socially optimal level of homophily to the equilibrium level. We show that even though the equilibrium level of homophily can be high in absolute terms, there may be too little homophily in equilibrium relative to the social optimal level:

**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

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.<sup>13</sup> 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.

A policy maker can also influence social welfare by strengthening cultural identity, for example by subsidizing cultural programs. Strengthening players' cultural identity has a direct impact on coordination payoffs by reducing strategic uncertainty. There is also an indirect effect, through the adjustment of the level of homophily in response to changes in cultural identity. This further reduces strategic uncertainty. The next result shows that the overall effect on social welfare is positive.

**Corollary 4.5.** *Suppose payoffs are fixed (i.e.,  $\alpha = \infty$ ). Policies that strengthen cultural identity lead to more homophily and improve welfare.*

These results may explain the popularity of policies that aim to strengthen cultural identity. For example, a strong organizational culture is often described as a key to a company's success (e.g., [Tichy and Sherman, 2005](#)). In 19th-century Europe, newly formed nation states built national museums to strengthen national identity ([Macdonald, 2012](#)). Social movements in 19th-century U.S. stimulated public school enrollment to build a new, common identity ([Meyer et al., 1979](#)). While such narratives are both widespread and intuitive, they are difficult to formalize within the standard framework. First, the standard framework does not explicitly model cultural identity. If we bring cultural identity into the standard framework by augmenting the standard framework with a signal structure, then the welfare implications of strengthening players' cultural identity may be ambiguous. This is because there are multiple (correlated) equilibria. Since any introspective equilibrium is a correlated equilibrium, there is a correlated equilibrium where strengthening cultural identity improves welfare. But, strengthening cultural identity also changes the set of equilibria, and there are instances where strengthening players' cultural identity gives rise to new equilibria with lower welfare.

---

<sup>13</sup>A player's choice also affects the payoffs of members of the other group. These effects go in the same direction.

## 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. For example, following a technological innovation, social interaction patterns may not readily adjust. So, a policy maker may have to choose a policy before it is fully known what the benefits of the innovation is when it is adopted by the public. So, we now consider the case the degree of economic stability  $\alpha$  is close to its minimum value, that is, close to 1. As before, this is a useful benchmark case that allows us to emphasize the driving forces. To gain more insight, it is 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  $\alpha$  decreases to 1.*

So, the level of homophily that maximizes total coordination payoffs decreases with the strength of cultural identity. In direct contrast, in stable economic environments, 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. To achieve the high payoff  $v^*$ , players may have to go against their instincts, when their initial impulse is to choose the other action. The main obstacle is the risk of miscoordinating. This risk is smaller for the minority because they are less likely to meet a similar opponent and so their initial impulse is a weaker signal. So, if  $v^*$  can be moderately high then the minority may decide to play the Pareto superior action. But this is not sufficient to induce the majority to risk miscoordination, unless the size of the minority is sufficiently large. Thus, it may pay to weaken cultural identity so that homophily is reduced and the size of the minority is increased. This may lead the majority to follow the minority and choose the Pareto superior action. This intuition 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.

The central point in these results is that optimal policy will tend to increase homophily in stable environments and to decrease in unstable ones. This holds when the policy maker cares not only about coordination payoffs. In the case that project payoffs are also taken into account, 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  $h^\alpha$  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. If cultural identity is strong (i.e.,  $Q > Q^*$ ), there is a strong pressure to conform. Any level of homophily that deviates from the 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 maximize coordination payoffs. On the other hand, if players' cultural identity is weak (i.e.,  $Q < Q^*$ ), then the pressure to conform is weak, for the minority or the majority. So, the level of homophily should lower than  $h_C$  to induce players to choose the project they intrinsically prefer.

Proposition 4.7 suggests that answers to policy questions such as whether students should be free to select roommates or should have roommates assigned at random can be subtle. In some cases the utility gains from freedom of choice dominates and, in others, it does not. Analogously, in schools that 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 restrict choice (Boisjoly et al., 2006; Burns et al., 2013; Sacerdote and Marmaros, 2006).

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  $v$ .*

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 determine 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.*

This result is again in a marked contrast with the results for stable economic environments. In a stable economic environment, strengthening cultural identity enhances social welfare, in uncertain economic environments, strengthening cultural identity decreases social welfare. The key difference is that in stable economic environment, greater cultural identity leads to successful coordination because it reduces strategic uncertainty. However, in an uncertain economic environment this can be harmful, as it increases the pressure to conform. So, in an uncertain environment, the social optimally requires a proper balance where the minority is small enough to face limited pressure to conform, yet large enough so that it can influence the majority.

Once again, our results are consistent with common narratives on the trade-offs involving social policy, but standard equilibrium analysis does produces these results. 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 segregated societies may find it harder to break out of low-payoff equilibria than more open-minded societies (Mokyr, 1990), and how this interacts with the social structure.

## 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.<sup>14</sup> 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

---

<sup>14</sup>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.

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  $j, \ell$  have chosen the same project  $\pi = a, b$ , and invest effort  $e_j$  and  $e_\ell$ , respectively, then the probability that they are matched (and play the coordination game) is

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

where  $E^\pi$  is the total effort of the players with project  $\pi$ .<sup>15</sup> 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  $E^\pi$ , we ensure that the network does not become arbitrarily dense as the number of players grows large.<sup>16</sup> 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.<sup>17</sup>

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

**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.,  $h > \varepsilon$ ).*

---

<sup>15</sup>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.

<sup>16</sup>See, e.g., Cabrales et al. (2011) and Galeotti and Merlino (2014) for applications of this model in economics.

<sup>17</sup>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).

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.

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.,  $|N| \rightarrow \infty$ ). 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*

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

*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:<sup>18</sup>

**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

---

<sup>18</sup>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  $e^- < e^*$ .

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, however, is needed, to establish the extent that these observations can be attributed to the economic and cultural factors related to strategic uncertainty.

## 6. Conclusions

This paper introduces a novel approach to model players' introspective process, grounded in the Theory of Mind. We show that high levels of homophily are possible even if there are no group-specific externalities and no direct preference for interacting with similar players. Modeling players' introspective process explicitly makes it possible to derive unique predictions and 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 segregation can be optimal in a stable economic environment, diversity and integration is better in 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 the extent to which beliefs and the need 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 this direction.

## 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  $b$ . Likewise, on average, players in group  $B$  have an intrinsic preference for  $b$ .

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

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

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

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  $b$ .

## B. Equilibrium analysis

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

We study the correlated equilibria of the extended game: in the first stage, players choose a project and are matched with players with the same project; and in the second stage, players play the coordination game with their partner. It is not hard to see that every introspective

equilibrium is a correlated equilibrium. The game has more equilibria, though, even if we fix the signal structure. For example, in the coordination stage, the strategy profile under which all players choose the same fixed action regardless of their signal is a correlated equilibrium, as is the strategy profile under which half of the players in each group choose  $s^1$  and the other half of the players choose  $s^2$ , or where players go against the action prescribed by their signal (i.e., choose  $s^2$  if and only the signal is  $s^1$ ). Given this, there is a plethora of equilibria for the extended game.

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

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

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

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

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

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

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

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

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  $\beta$ , are so large that they dominate any intrinsic preference.

But even if  $\beta$  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  $\varepsilon$  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.

The multiplicity of equilibria in the standard setting makes it difficult to derive unambiguous comparative statics and clear welfare implications. This is because as parameters are adjusted, the set of equilibria changes. Consider, for example, the effect of strengthening players' cultural identity. As any introspective equilibrium is a correlated equilibrium, there is a correlated equilibrium where strengthening cultural identity leads to higher levels of homophily (Corollary 3.3) and improves social welfare (Corollary 4.5). But, strengthening cultural identity also changes the set of equilibria, and in some cases, strengthening players' cultural identity gives rise to new equilibria with lower levels of homophily, and reduce social welfare.

## 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.<sup>19</sup> Each player is matched exactly once.<sup>20</sup> 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

---

<sup>19</sup>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.

<sup>20</sup>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.

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*

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

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

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

*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:

**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.<sup>21</sup>

## 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:

---

<sup>21</sup>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.

	$s^1$	$s^2$		$s^1$	$s^2$
$s^1$	$v, v$	$0, 0$	$s^1$	$V, V$	$0, 0$
$s^2$	$0, 0$	$v, v$	$s^2$	$0, 0$	$V, V$
	Own group			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

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

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

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

Note that  $\beta_{CS}$  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 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:<sup>22</sup>

**Lemma D.1.** *Suppose skill complementarities dominate and that  $\beta_{CS} > -\frac{1}{2}$ . The limit  $p^\pi$  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  $\pi$ .

---

<sup>22</sup>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.

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  $p$ .  $\square$

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

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

It can be checked that  $p$  is increasing in  $\beta_{CS}$ , 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:*

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

*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  $\tilde{C}^\infty(p)$  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  $\varepsilon$ , 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.

## E. Proofs for Sections 2–3

### 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

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

where  $\mathbb{P}(m | i)$  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  $\mathbb{P}(m) = \frac{1}{2}$ ,  $\mathbb{P}(i | i) = q^2 + (1 - q)^2$  and  $\mathbb{P}(j | i) = 1 - q^2 - (1 - q)^2$ , 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

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

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

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

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  $\pi$ .

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  $1 - p_{k-1}^a$ . Applying Proposition 2.1 (with  $\hat{p} = p_{k-1}^a$  and  $\hat{p} = 1 - p_{k-1}^a$ ) 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

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

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

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

where we have used the expression for the tail distribution  $H_\varepsilon$  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  $\pi$ . 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

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

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  $(2p_0^\pi - 1) \cdot \beta \leq 1 - 2\varepsilon$  and  $1 - 2\varepsilon \geq 0$ . 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  $p_{k-1}^\pi \geq p_{k-2}^\pi$  and that  $p_{k-1}^\pi \in (\frac{1}{2}, 1]$ . By a similar argument as above,

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

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

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

or, equivalently,

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

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

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

So, it suffices to show that

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

The above inequality holds if and only if

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

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

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  $p^\pi$  exists. The argument clearly does not depend on the project  $\pi$ , 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  $p_k \geq p_{k-1}$  for all  $k$ . By the monotone sequence convergence theorem,  $p = \sup_k p_k$ , and by the inductive argument,  $p \in (\frac{1}{2} + \varepsilon, 1]$ . It is easy to see that  $p = 1$  if and only if  $H_\varepsilon(-(2 \cdot 1 - 1) \cdot \beta) = 1$ , which holds if and only if  $\beta \geq 1 - 2\varepsilon$ .

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

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

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

$$p = 1 - \frac{1}{2x} \cdot (x - (2p - 1) \cdot \beta)^2. \quad (\text{E.2})$$

Equation (E.2) has two roots,

$$r_1 = \frac{1}{2} + \frac{1}{4\beta^2} \left( (2\beta - 1) \cdot x + \sqrt{4\beta^2 x - (4\beta - 1) \cdot x^2} \right)$$

and

$$r_2 = \frac{1}{2} + \frac{1}{4\beta^2} \left( (2\beta - 1) \cdot x - \sqrt{4\beta^2 x - (4\beta - 1) \cdot x^2} \right).$$

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 3.3, it is straightforward to verify that the derivative of  $p$  with respect to  $\beta$  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).

## F. Proofs for Section 4

### F.1. Preliminary results

Fix an assignment of players to project (i.e., a mapping from the set of players in each group to the set of projects), and let  $p$  be the corresponding probability that a player with the group-preferred project (e.g., an  $A$ -player assigned to project  $a$ ) interacts with a member of his own group; note that a player with the non-group preferred project (e.g., a  $B$ -player assigned to project  $a$ ) has a probability of  $1 - p$  of interacting with a member of his own group. Note that in any social optimum, at least half the players of each project are assigned to the group-preferred project (i.e.,  $p \geq \frac{1}{2}$ ).<sup>23</sup> Recall that action  $s^1$  gives a payoff of  $v^*$  if both players choose it, while action  $s^2$  gives a payoff equal to  $v \leq v^*$ ; we refer to  $s^1$  and  $s^2$  as the Pareto superior and the Pareto inferior action.

#### F.1.1. Introspective equilibrium

We characterize the introspective equilibrium in the coordination game when one action is potentially Pareto superior. The project assignments (and thus the probability  $p$ ) and the payoffs are commonly known among the players. At level 0, each player follows his impulse in the coordination game, as before. For  $k > 1$ , suppose that all players follow their impulse

---

<sup>23</sup>To see this, suppose by contradiction that a share  $p < \frac{1}{2}$  of players (of a given group, say  $A$ ) is assigned to 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 is assigned to that project (and the other players to 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 is assigned to the project that they intrinsically prefer.

at level  $k - 1$ . At level  $k$ , an  $A$ -player assigned to project  $a$  with value  $w_j^{A,a}$  for project  $a$  and the impulse to play the Pareto superior action  $s^1$  chooses  $s^1$  if and only if<sup>24</sup>

$$v^* \cdot [p \cdot Q + (1 - p) \cdot \tfrac{1}{2}] + w_j^{A,a} \geq v \cdot [p \cdot (1 - Q) + (1 - p) \cdot \tfrac{1}{2}] + w_j^{A,a},$$

or, equivalently, if and only if

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

Since the left-hand side is at least 1, and the right-hand side is less than 1 (as  $Q > \tfrac{1}{2}$ ), this inequality is always satisfied. Likewise, a  $B$ -player assigned to project  $a$  who has the impulse to play  $s^1$  chooses  $s^1$  if and only if

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

Again, this inequality is satisfied for any combination of parameters given that  $Q > \tfrac{1}{2}$  and  $v^* \geq v$ . So, if all players follow their impulse at level  $k - 1$ , a player assigned to project  $a$  with an impulse to play  $s^1$  chooses  $s^1$ ; the analogous statement clearly holds for players assigned to project  $b$ . We next consider players that have an impulse to play the Pareto inferior action  $s^2$ . An  $A$ -player assigned to project  $a$  with value  $w_j^{A,a}$  for project  $a$  and the impulse to play action  $s^2$  goes against his impulse and chooses  $s^1$  if and only if<sup>25</sup>

$$v^* \cdot [p \cdot (1 - Q) + (1 - p) \cdot \tfrac{1}{2}] + w_j^{A,a} > v \cdot [p \cdot Q + (1 - p) \cdot \tfrac{1}{2}] + w_j^{A,a},$$

or, equivalently, if and only if

$$\frac{v^*}{v} > \frac{p \cdot Q + (1 - p) \cdot \tfrac{1}{2}}{p \cdot (1 - Q) + (1 - p) \cdot \tfrac{1}{2}} =: T_D(p, Q),$$

where  $D$  stands for “dominant group.” Likewise, a  $B$ -player assigned to project  $a$  who has the impulse to play  $s^2$  goes against his impulse and chooses  $s^1$  if and only if

$$\frac{v^*}{v} > \frac{(1 - p) \cdot Q + p \cdot \tfrac{1}{2}}{(1 - p) \cdot (1 - Q) + p \cdot \tfrac{1}{2}} =: T_M(p, Q),$$

where  $M$  stands for “minority group.” This characterizes the behavior of players assigned to project  $a$  that have an impulse to play the Pareto inferior action; of course, the characterization of the behavior of players assigned to project  $b$  that have an impulse to play the Pareto inferior action is similar. This gives the following result:

---

<sup>24</sup>As before, we assume that a player follows his impulse if he is indifferent. This does not affect our results.

<sup>25</sup>As before, we assume that a player follows his impulse if he is indifferent. This does not affect our results.

**Lemma F.1.** *If all players follow their impulse at level  $k - 1$ , then at level  $k$ :*

- (a) *a player with the group-preferred project chooses the Pareto superior action, regardless of his impulse, if and only if*

$$\frac{v^*}{v} > T_D(p, Q),$$

*and follows his impulse otherwise;*

- (b) *a player with the non-group preferred project chooses the Pareto superior action, regardless of his impulse, if and only if*

$$\frac{v^*}{v} > T_M(p, Q),$$

*and follows his impulse otherwise.*

Since  $Q > \frac{1}{2}$ , we have  $T_M(p, Q) \leq T_D(p, Q)$  for all  $p \geq \frac{1}{2}$ . So, if at level  $k - 1$ , all players follow their impulse, then at level  $k$ :

- if  $\frac{v^*}{v} \geq T_D(p, Q)$ , all players choose the Pareto superior action, regardless of their impulse;
- if  $\frac{v^*}{v} \in [T_M(p, Q), T_D(p, Q))$ , players assigned to the non-group preferred project choose the Pareto superior action, regardless of their impulse, while players assigned to the group-preferred project follow their impulse;
- if  $\frac{v^*}{v} < T_M(p, Q)$ , all players follow their impulse.

It can easily be checked that if at level  $k$ , all players choose the Pareto superior action (regardless of their impulse), then at level  $k + 1$ , all players choose the Pareto superior action. Also, if at level  $k \geq 1$ , all players follow their impulse, then all players follow their impulse at level  $k + 1$  (since  $T_D(p, Q)$  and  $T_M(p, Q)$  are independent of players' level).

So, it remains to consider the case where players with the non-group preferred project (e.g.,  $B$ -players with project  $a$ ) choose the Pareto superior action and players with the group-preferred project (e.g.,  $A$ -players with project  $a$ ) follow their signal. Suppose that at level  $k \geq 1$ , players with the non-group preferred project choose the Pareto superior action, while players with the group-preferred project follow their signal. As before, at level  $k + 1$ , a player with an impulse to choose the Pareto superior action chooses that action. A player with the non-group preferred project and the impulse to choose the Pareto inferior action goes against his impulse and chooses the Pareto superior action at level  $k + 1$  if and only if

$$v^* \cdot \left[ (1 - p) \cdot 1 + p \cdot \frac{1}{2} \right] > v \cdot \left[ (1 - p) \cdot 0 + p \cdot \frac{1}{2} \right],$$

a condition that is satisfied whenever  $p < 1$ , where the factor 1 stems from the fact that all players with the non-group preferred project choose the Pareto superior action at level  $k$ , regardless of their impulse. Since there are players with the non-group preferred project only if  $p < 1$ , we will largely ignore this condition in what follows.

A player with the group-preferred project and the impulse to choose the Pareto inferior action goes against his impulse and chooses the Pareto superior action at level  $k + 1$  if and only if

$$v^* \cdot [(1 - p) \cdot 1 + p \cdot Q] > v \cdot [(1 - p) \cdot 0 + p \cdot (1 - Q)],$$

or, equivalently, if and only if

$$\frac{v^*}{v} > \frac{pQ}{1 - pQ} =: T_C(p, Q).$$

This gives the following result:

**Lemma F.2.** *If at level  $k \geq 1$ , players with the non-group preferred project choose the Pareto superior action, while players with the group-preferred project follow their signal. At level  $k+1$ , players with the non-group preferred project choose the Pareto superior action (if  $p < 1$ ); and a player with the group-preferred project chooses the Pareto superior action, regardless of his impulse, if and only if*

$$\frac{v^*}{v} > T_C(p, Q),$$

*and follows his impulse otherwise.*

For given  $p$  and  $Q$ , either  $T_M(p, Q) \geq T_C(p, Q)$  or  $T_M(p, Q) < T_C(p, Q)$ . In the remainder, we write  $T_C$  and  $T_M$  for  $T_C(p, Q)$  and  $T_M(p, Q)$ , respectively, for simplicity. Using Lemmas F.1 and F.2, we can prove the following result.

**Proposition F.3.** *For any  $p$  and  $Q$ , there is a unique equilibrium for the coordination game.*

(a) *Suppose  $T_M \geq T_C$ . Then:*

- *If  $\frac{v^*}{v} > T_M$ , all players choose the Pareto superior action, regardless of their impulse;*
- *If  $\frac{v^*}{v} \leq T_M$ , all players follow their impulse.*

(b) *Suppose  $T_M < T_C$ . Then:*

- *If  $\frac{v^*}{v} > T_C$ , all players choose the Pareto superior action, regardless of their impulse;*
- *If  $\frac{v^*}{v} \in (T_M, T_C]$ , all players with the non-group preferred project choose the Pareto superior action, and all players with the group preferred project follow their impulse;*
- *If  $\frac{v^*}{v} \leq T_M$ , all players follow their impulse.*

### F.1.2. Coordination payoffs

Proposition F.3 allows us to calculate the total (expected) coordination payoff as a function of  $p$ . It will be convenient to work with payoffs per project. Define  $\tilde{C}^\alpha(p)$  to be the total (expected) coordination payoff attained in the unique introspective equilibrium by players with project  $a$  when a share  $p$  of players choose the group-preferred project. By symmetry,  $\tilde{C}^\alpha(p)$  is equal to the total expected coordination payoff for players with project  $b$ . So,  $C^\alpha(h) = 2\tilde{C}^\alpha(p)$  when the level of homophily is  $h = p - \frac{1}{2}$ .

Proposition F.3 can be used to calculate the expected coordination payoffs. If  $T_M < T_C$ , then

$$\begin{aligned} \tilde{C}^\alpha(p) = & \mathbb{P}\left(\frac{v^*}{v} \geq T_C\right) \mathbb{E}[v^* \mid \frac{v^*}{v} \geq T_C] + \\ & \mathbb{P}\left(\frac{v^*}{v} \in (T_M, T_C)\right) \cdot (1-p) \cdot \mathbb{E}[v^* \mid \frac{v^*}{v} \in (T_M, T_C)] + \\ & \mathbb{P}\left(\frac{v^*}{v} \in (T_M, T_C)\right) \cdot p^2 \cdot Q \cdot \left[\frac{v}{2} + \frac{1}{2} \cdot \mathbb{E}[v^* \mid \frac{v^*}{v} \in (T_M, T_C)]\right] + \\ & \mathbb{P}\left(\frac{v^*}{v} \leq T_M\right) \cdot ((p^2 + (1-p)^2) \cdot Q + p \cdot (1-p)) \cdot \left[\frac{v}{2} + \frac{1}{2} \cdot \mathbb{E}[v^* \mid \frac{v^*}{v} \leq T_M]\right]. \end{aligned} \quad (\text{F.1})$$

The first term is the expected coordination payoff if the realization of ratio  $\frac{v^*}{v}$  is sufficiently high so that in the unique introspective equilibrium, all players choose the Pareto superior action. The second and the third term are the expected payoffs for the players with the non-group preferred project and with the group-preferred project, respectively, if  $\frac{v^*}{v}$  is such that players with the non-group preferred players choose the Pareto superior action, and the other players follow their impulse. The last term is the expected coordination payoff if  $\frac{v^*}{v}$  is such that in the unique introspective equilibrium, all players follow their impulse. If  $T_M \geq T_C$ , the per-project expected coordination payoff is

$$\begin{aligned} \tilde{C}^\alpha(p) = & \mathbb{P}\left(\frac{v^*}{v} \geq T_M\right) \mathbb{E}[v^* \mid \frac{v^*}{v} \geq T_M] + \\ & \mathbb{P}\left(\frac{v^*}{v} \leq T_M\right) \cdot ((p^2 + (1-p)^2) \cdot Q + p \cdot (1-p)) \cdot \left[\frac{v}{2} + \frac{1}{2} \cdot \mathbb{E}[v^* \mid \frac{v^*}{v} \leq T_M]\right]. \end{aligned} \quad (\text{F.2})$$

Again, the first term is the expected coordination payoff if the realization of  $\frac{v^*}{v}$  is sufficiently high so that all players choose the Pareto superior action in the unique introspective equilibrium, whereas the last term is the expected coordination payoff if  $\frac{v^*}{v}$  is low and all players follow their impulse in the unique introspective equilibrium.

### F.1.3. Project values

Let  $\tilde{\Pi}(p)$  be the total project value for players that are assigned to project  $a$  when a share  $p$  of players with the strongest intrinsic preference for the group-preferred project are assigned to that project. That is,  $\tilde{\Pi}(p)$  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  $a$ . As before,  $\tilde{\Pi}(p)$  is equal to the total value derived from project  $b$ . So,  $\Pi(h) = 2\tilde{\Pi}(p)$  when the level of homophily is  $h = p - \frac{1}{2}$ . The next result characterizes the payoffs  $\tilde{\Pi}(p)$  derived from a project.

**Lemma F.4.** *In any social optimum where the share of players assigned to the group-preferred project is  $p \geq \frac{1}{2}$ , the total value derived from a project is*

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

**Proof.** To calculate total project value  $\tilde{\Pi}(p)$ , fix a group, say  $A$ . Recall that in any social optimum, the share  $p$  of players of group  $A$  with the strongest intrinsic preference for project  $a$  is assigned to project  $a$ . So, in any social optimum, all  $A$ -players for whom the difference  $w_j^{A,a} - w_j^{A,b}$  exceeds a certain threshold  $y$  is assigned to project  $a$ , and the other  $A$ -players are assigned to 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 assigned to 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,

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

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

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

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 are assigned to project  $b$  is given by

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

The second term is for  $A$ -players for whom  $w_j^{A,a}$  is so small (relative to the threshold  $y$ ) that they are assigned to project  $b$  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  $\tilde{\Pi}(p)$  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  $z = z(p)$  are assigned to project  $a$  (and the other  $A$ -players are assigned to 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

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

It will be convenient to work with a nonnegative threshold, so define  $y := -z \geq 0$ . Then, rewriting gives<sup>26</sup>

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

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

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

where the first term is for  $A$ -players for whom  $w_j^{A,b}$  is sufficiently low that they are assigned to 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  $\tilde{\Pi}(p)$  for  $p \in [\frac{1}{2} + \varepsilon, 1]$ .  $\square$

We are now ready to derive an expression for (expected) social welfare as a function of  $p \in [\frac{1}{2}, 1]$ . Social welfare per project is

$$\tilde{W}^\alpha(p) = \tilde{C}^\alpha(p) + \tilde{\Pi}(p),$$

and social welfare (over all projects) as a function of the level of homophily  $h = p - \frac{1}{2}$  is

$$W^\alpha(h) = 2\tilde{W}^\alpha(p). \tag{F.3}$$

We will use this expression in the following to characterize the socially optimal level of homophily in stable and uncertain economic environments. The following result will also be useful:

---

<sup>26</sup>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.

**Lemma F.5.** *Suppose that payoffs are fixed (i.e.,  $\alpha = \infty$ ), and that a share  $p \geq \frac{1}{2}$  of players in each group is assigned to the group-preferred project. The per-project coordination payoff is*

$$\tilde{C}^\infty(p) := v \cdot [Q \cdot (p^2 + (1-p)^2) + 2 \cdot p \cdot (1-p) \cdot \frac{1}{2}]. \quad (\text{F.4})$$

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

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

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

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

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

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

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

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

Adding all terms together gives  $\tilde{C}^\infty(p)$ . A similar calculation, of course, applies to project  $b$ .  $\square$

## F.2. Proof of Proposition 4.1

We use the notation and results from Appendix F.1. Fix  $Q$  and  $v$ . From Lemmas F.1 and F.2, it follows that conditional on  $v^*$ , per-project social welfare in the unique introspective equilibrium is

$$\mathcal{W}^\alpha(p) = \begin{cases} v^* + \tilde{\Pi}(p) & \text{if } \frac{v^*}{v} > \max\{T_M, T_C\}; \\ p^2 \cdot Q \cdot (\frac{v}{2} + \frac{v^*}{2}) + (1-p) \cdot v^* + \tilde{\Pi}(p) & \text{if } \frac{v^*}{v} \in (T_M, T_C]; \\ ((p^2 + (1-p)^2) \cdot Q + p \cdot (1-p)) \cdot (\frac{v}{2} + \frac{v^*}{2}) + \tilde{\Pi}(p) & \text{if } \frac{v^*}{v} \leq T_M; \end{cases}$$

where we write  $T_C$  and  $T_M$  for  $T_C(p, Q)$  and  $T_M(p, Q)$ , respectively, as in Appendix F.1. It is straightforward to verify that  $\mathcal{W}^\alpha(\cdot)$  is continuous and bounded, and that  $\mathcal{W}^\alpha(p)$  increases with  $\frac{v^*}{v}$  for every  $\alpha, p, Q$ . Moreover, for  $\alpha' > \alpha$ , the cumulative distribution function  $F^\alpha(y) := 1 - y^{-\alpha}$  of  $\frac{v^*}{v}$  first-order stochastically dominates  $F^{\alpha'}(y)$ . It follows that the expectation  $\tilde{W}^\alpha(p)$  of  $\mathcal{W}^\alpha(p)$  (under  $F^\alpha$ ) increases monotonically in  $\alpha$  (for every  $p$ ). Using Lemma F.5, we can

define the per-project social welfare when the payoff to both actions is  $v$  (with probability 1), given by

$$\widetilde{W}^\infty(p) := ((p^2 + (1-p)^2) \cdot Q + p \cdot (1-p)) \cdot v + \widetilde{\Pi}(p)$$

where  $\widetilde{\Pi}(p)$  is as defined previously. Of course, the per-project social welfare  $\widetilde{W}^\infty(p)$  when payoffs are fixed is equal to the expectation of  $\mathcal{W}^\alpha(p)$  (for given  $p$ ) under the degenerate distribution  $F^\infty$  that puts probability 1 on  $\frac{v^*}{v} = 1$ . By standard arguments, we have  $\widetilde{W}^\alpha(p) \rightarrow \widetilde{W}^\infty(p)$  for every  $p$  as  $\alpha$  goes to infinity.

As  $\widetilde{W}^\alpha(\cdot)$  and  $\widetilde{W}^\infty(\cdot)$  are continuous functions of  $p \in [\frac{1}{2}, 1]$ , we can apply Dini's theorem to show that  $\widetilde{W}^\alpha$  converges uniformly to  $\widetilde{W}^\infty$  as  $\alpha \rightarrow \infty$ .

By Proposition 4.2 below,<sup>27</sup> when payoffs are fixed ( $\alpha = \infty$ ), per-project social welfare  $\widetilde{W}^\infty(\cdot)$  has a unique maximizer  $p^\infty \in [\frac{1}{2}, 1]$ .

The result now follows from the following lemma:

**Lemma F.6.** *For  $\alpha > 1$ , let  $p^\alpha$  be the probability that maximizes per-project social welfare  $\widetilde{W}^\alpha(p)$ . Then  $p^\alpha \rightarrow p^\infty$  as  $\alpha$  goes to infinity.*

**Proof.** Let  $\eta > 0$ , and define

$$P_\eta := \{p \in [\frac{1}{2}, 1] : |p - p^\infty| < \eta\}.$$

Since  $P_\eta$  is open (in the usual topology), the intersection  $[\frac{1}{2}, 1] \cap P_\eta^c$  is compact, where we have used the standard notation  $A^c$  to denote the complement of a set  $A$ . Since  $\widetilde{W}^\infty(\cdot)$  is continuous,  $\widetilde{W}^\infty(\cdot)$  attains its supremum  $p^*$  on  $[\frac{1}{2}, 1] \cap P_\eta^c$ . Since  $p^\infty$  is the unique maximizer of  $\widetilde{W}^\infty(\cdot)$ , we have

$$\widetilde{W}^\infty(p^\infty) - \widetilde{W}^\infty(p^*) =: \delta > 0$$

(recall that  $p^* \notin P_\eta$ , so  $p^* \neq p^\infty$  by definition). As  $\widetilde{W}^\alpha(\cdot)$  converges uniformly to  $\widetilde{W}^\infty(\cdot)$  as  $\alpha$  goes to infinity, there is  $\underline{\alpha}$  such that for all  $\alpha > \underline{\alpha}$ ,

$$\sup_{p \in [\frac{1}{2}, 1] \cap P_\eta^c} |\widetilde{W}^\alpha(p) - \widetilde{W}^\infty(p)| < \frac{\delta}{2}.$$

Hence, for all  $p \in P_\eta^c$  and  $\alpha > \underline{\alpha}$ ,

$$\begin{aligned} \widetilde{W}^\alpha(p) &< \widetilde{W}^\infty(p) + \frac{\delta}{2} \\ &\leq \widetilde{W}^\infty(p^*) + \frac{\delta}{2} \\ &= \widetilde{W}^\infty(p^\infty) - \frac{\delta}{2}, \end{aligned} \tag{F.5}$$

---

<sup>27</sup>Note that Proposition 4.2 does not rely on the present result.

where the last equality follows from the definition of  $\delta$ . Also, by uniform convergence, there is  $\underline{\alpha}' > 1$  such that for all  $\alpha > \underline{\alpha}'$ ,

$$\sup_{p \in P_\eta} |\widetilde{W}^\alpha(p) - \widetilde{W}^\infty(p)| < \frac{\delta}{2}.$$

It follows that for all  $p \in P_\eta$  and  $\alpha > \underline{\alpha}'$ ,

$$\widetilde{W}^\alpha(p) > \widetilde{W}^\infty(p) - \frac{\delta}{2}.$$

In particular,

$$\widetilde{W}^\alpha(p^\infty) > \widetilde{W}^\infty(p^\infty) - \frac{\delta}{2}. \quad (\text{F.6})$$

It follows from (F.5) and (F.6) that there is  $\underline{\alpha}_\eta > 1$  such that for all  $\alpha > \underline{\alpha}_\eta$ , the probability  $p^\alpha$  that maximizes  $\widetilde{W}^\alpha(p)$  is in  $P_\eta$ .  $\square$

It follows from Lemma F.6, as  $\alpha$  goes to infinity, the level of homophily  $h^\alpha$  that maximizes social welfare  $W^\alpha(h)$  converges to the level of homophily  $h^\infty$  that maximizes social welfare  $W^\infty(h)$  when payoffs are fixed. That social welfare also converges follows from the continuity of the social welfare function as well as the uniform convergence of  $\widetilde{W}^\alpha(\cdot)$  to  $\widetilde{W}^\infty(\cdot)$ .  $\square$

### F.3. Proof of Proposition 4.2

Maximizing social welfare is equivalent to maximizing per-project social welfare. By the results in Appendix F.1, per-project welfare as a function of the probability  $p$  is given by

$$\widetilde{W}^\infty(p) = v \cdot [Q \cdot (p^2 + (1-p)^2) + 2 \cdot p \cdot (1-p) \cdot \frac{1}{2}] + \widetilde{\Pi}(p),$$

where the first term are the per-project coordination payoffs (Lemma F.5), and the second term is the value derived from a project, as characterized in Lemma F.4.

It will be useful to define the *marginal benefit of interacting with the own group* as

$$\beta := Qv - \frac{1}{2}v.$$

The first term is the expected coordination payoff for a player of interacting with a member of the own group; the second term is the expected coordination payoff of interacting with a member of the other group. So,  $\beta$  is the marginal benefit to a player when the probability of interacting with the own group is increased. Since  $Q > \frac{1}{2}$ , we have  $\beta > 0$ . We characterize the socially optimal level of homophily both for the case that the marginal benefit of interacting with the own group  $\beta$  is positive, as well as for the case that  $\beta$  is negative. The characterization for this latter case will be useful when we consider complementarities of skills between groups in Appendix D.

As in the proof of Lemma F.4, we need to consider two cases.

**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

$$2 \cdot (2p - 1)\beta + \frac{1}{2x} \left[ 4x(1 - p - \frac{x}{2}) - (1 - p - \frac{x}{2})^2 \right].$$

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

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

and

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

It is straightforward to verify that  $r_2 \leq \frac{1}{2}$  whenever  $x \geq \frac{1}{9}$ . Also, if  $x \geq \frac{1}{9}$ , 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:

$$r'_1 = \frac{1}{2} + \frac{x}{4\beta^2} \left[ \beta - \frac{1}{4} + \sqrt{\frac{\beta^2}{x} - \frac{\beta}{2} + \frac{1}{16}} \right],$$

and

$$r'_2 = \frac{1}{2} + \frac{x}{4\beta^2} \left[ \beta - \frac{1}{4} - \sqrt{\frac{\beta^2}{x} - \frac{\beta}{2} + \frac{1}{16}} \right].$$

For any combination of parameters,  $r'_2 \leq \frac{1}{2}$ . Clearly,  $r'_1 > r'_2$ ; moreover,  $r'_1$  is a saddle point (and thus a point of inflection) if and only if  $2\beta \geq x$ . If  $2\beta \geq x$ , 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  $h^* = r'_1 - \frac{1}{2} > \varepsilon$ .  $\square$

## F.4. Proof of Proposition 4.6

Maximizing the total coordination payoff  $C^\alpha(h)$  is equivalent to maximizing the per-project expected coordination payoff  $\tilde{C}^\alpha(p)$  (where  $p = h + \frac{1}{2}$ ). Note that the threshold values  $T_M(p, Q)$  and  $T_C(p, Q)$  defined in Appendix F.1 are (strictly) decreasing and increasing in  $p$ , respectively (for a given  $Q$ ). It will be useful to define  $p_c$  to be the probability  $p$  for which  $T_C(p, Q) = T_M(p, Q)$  (for given  $Q$ ), that is,

$$p_c := \frac{Q}{2Q - \frac{1}{2}},$$

so that  $T_C(p, Q) \geq T_M(p, Q)$  if and only if  $p \geq p_c$ . Note that for  $\varepsilon$  sufficiently small,  $p_c > \frac{1}{2} + \varepsilon$  for all  $Q \in [\frac{1}{2}, 1]$ . The associated level of homophily is

$$h_c := p_c - \frac{1}{2} = \frac{Q}{8Q - 2}.$$

We consider two cases.

**Case 1:**  $p \leq p_c$  ( $T_M \geq T_C$ ). In the unique introspective equilibrium, either all players follow their impulse (if the realization of  $\frac{v^*}{v}$  is less than  $T_M$ ), or all players choose the Pareto superior action (if  $\frac{v^*}{v} \geq T_M$ ). By (F.2), and using that  $\frac{v^*}{v}$  is Pareto distributed with parameter  $\alpha$ , we have

$$\begin{aligned} \tilde{C}^\alpha(p) = & \frac{\alpha v}{\alpha - 1} T_M^{-\alpha+1} + \\ & \frac{v}{2} \cdot ((p^2 + (1-p)^2) \cdot Q + p \cdot (1-p)) \cdot \left(1 - T_M^{-\alpha} + \frac{\alpha}{\alpha - 1} (1 - T_M^{-\alpha+1})\right). \end{aligned}$$

Differentiating this expression shows that  $\tilde{C}^\alpha(\cdot)$  increases with  $p$  for every  $Q \in [\frac{1}{2}, 1]$ .

**Case 2:**  $p > p_c$  ( $T_M < T_C$ ). In the unique introspective equilibrium, there are three possibilities, depending on the realization of  $\frac{v^*}{v}$ . If the realization of ratio  $\frac{v^*}{v}$  is sufficiently high (viz.,  $\frac{v^*}{v} > T_C$ ), then in the unique introspective equilibrium, all players choose the Pareto superior action. If  $\frac{v^*}{v}$  is sufficiently low (viz.,  $\frac{v^*}{v} \leq T_M$ ), then in the unique introspective equilibrium, all players follow their impulse. If  $\frac{v^*}{v}$  is between these two extremes (viz.,  $\frac{v^*}{v} \in (T_M, T_C]$ ), then players with the non-group preferred players choose the Pareto superior action, and the other players follow their impulse. By (F.1), using that  $\frac{v^*}{v}$  is Pareto distributed with parameter  $\alpha$ , we can write

$$\begin{aligned} \frac{\tilde{C}^\alpha(p)}{v} = & \left(p - \frac{p^2 Q}{2}\right) \cdot \frac{\alpha T_C^{-\alpha+1}}{\alpha - 1} + \frac{\alpha}{\alpha - 1} \cdot \left(1 - p + \frac{p^2 Q}{2}\right) + \\ & \frac{\alpha}{\alpha - 1} \cdot \left((p^2 + (1-p)^2) \cdot \frac{Q}{2} + \frac{p}{2} \cdot (1-p) - (1-p) - \frac{p^2 Q}{2}\right) \cdot (1 - T_M^{-\alpha+1}) + \\ & \frac{p^2 Q}{2} \cdot (1 - T_C^{-\alpha}) + \left((p^2 + (1-p)^2) \cdot \frac{Q}{2} + \frac{p}{2} \cdot (1-p) - \frac{p^2 Q}{2}\right) \cdot (1 - T_M^{-\alpha}). \end{aligned}$$

□

## G. Proofs for Section 5

### G.1. 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  $N_0^{A,b}$  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<sup>28</sup>

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

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

---

<sup>28</sup>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.

$p_0 := \frac{1}{2} + \varepsilon > \frac{1}{2}$  (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  $g_1^M$ , respectively, then the result follows immediately from the fact that  $N_0^D$  first-order stochastically dominates  $N_0^M$ , as  $g_1^D$  is (strictly) increasing in  $N_0^D$  and (strictly) decreasing in  $N_0^M$ , and  $g_1^M$  is decreasing in  $N_0^D$  and increasing in  $N_0^M$  (again, strictly).

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

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

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

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

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

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

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

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

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  $N_1^{A,b}$  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  $N_1^{B,a}$  are clearly not independent (as  $N_1^{A,a} + N_1^{B,a} = N$ ),  $N_1^{A,a}$  and  $N_1^{B,b}$  are independent (and similarly if we replace  $N_1^{A,a}$ ,  $N_1^{B,a}$ , and  $N_1^{B,b}$  with  $N_1^{A,b}$ ,  $N_1^{B,a}$ , 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

$$e_k^D + e_k^M = \frac{v}{c}(Q + \frac{1}{2}). \quad (\text{G.1})$$

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

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

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,

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

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

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

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

$$e_{k-1}^D \leq \left(\frac{v}{c}\right) \cdot \mathbb{E} \left[ \frac{N_{k-1}^D \cdot Q + N_{k-1}^M \cdot \frac{1}{2}}{N_{k-1}^D + N_{k-1}^M} \right]. \quad (\text{G.3})$$

The result now follows by comparing Equations (G.2) and (G.3). Also, using that  $g_k^D$  is increasing and decreasing in its first and second argument, respectively, we have that

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

and it follows from (G.1) 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

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

Hence, the number  $N_k^{A,a}$  of  $A$ -players with project  $a$  (or, the number  $N_k^{B,b}$  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  $N_k^{A,b}$  of  $A$ -players with project  $b$  (or, the number  $N_k^{B,a}$  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 (G.1) 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  $e_1^D, e_2^D, \dots$  are monotone and bounded, so that by the monotone convergence theorem, their respective limits  $p := \lim_{k \rightarrow \infty} p_k$  and  $e^D := \lim_{k \rightarrow \infty} e_k^D$  exist (as does  $e^M := \lim_{k \rightarrow \infty} e_k^M = \frac{v}{c}(Q + \frac{1}{2}) - e^D$ ).  $\square$

## G.2. 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

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

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

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

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.<sup>29</sup> 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 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

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

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  $h^D(e^D)$ .<sup>30</sup>

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

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

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

<sup>29</sup>As before, the expectation is taken over  $N^D, N^M$  such that  $N^D > 0$  or  $N^M > 0$ .

<sup>30</sup>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  $n^D > 0, n^M = 0$ .

We can rewrite this condition as

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

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.

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.  $\square$

### G.3. 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  $h = p - \frac{1}{2}$ ).  $\square$

## 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.
- DeBruine, L. M. (2002). Facial resemblance enhances trust. *Proceeding of the Royal Society* 269, 1307–1312.
- 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.
- Guiso, L., P. Sapienza, and L. Zingales (2009). Cultural biases in economic exchange? *Quarterly Journal of Economics* 124, 1095–1131.
- 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.