A belief-based theory of homophily*

Willemien Kets[a,1,*], Alvaro Sandroni[b,2]

[a]Department of Economics, University of Oxford, Manor Road, Oxford OX13UQ, United Kingdom.
[b]Kellogg School of Management, Northwestern University, 2211 Campus Drive, Evanston, IL 60208, United States.

Abstract

Homophily, the tendency of people to associate with people similar to themselves, is a widespread phenomenon that has important economic consequences. We endogenize players' preferences for interacting with their own group by modeling the process by which players take others' perspective. Homophily emerges because players find it easier to put themselves into the shoes of members of their own group. The model sheds light on various empirical regularities and has novel welfare implications. In particular, policies that reduce homophily may not improve social welfare.

Keywords: Homophily, culture, theory of mind, strategic uncertainty, coordination

1. Introduction

Homophily, the tendency of people to interact with similar people, is a widespread phenomenon that has important economic consequences, affecting, for instance, investment in education [18], wages and employment [54], and the diffusion of information [29]. Much of the existing literature explains homophily by assuming a direct preference for associating with similar others [see 35, for a survey]. However, without a theory of the determinants of these preferences, it is hard to explain why homophily is observed in some cases, but not in others (beyond positing homophilous preferences only in the former settings). And, without a better understanding of the root causes of homophily, it is unclear whether policies aimed at reducing homophily improve social welfare.

We provide a theory of homophily that does not assume homophilous preferences. Rather, in our model, a preference to interact with similar others is a natural outcome of individuals' desire to reduce *strategic uncertainty*, i.e., uncertainty about others' actions. This approach allows us to shed light on various empirical phenomena that are difficult to explain with models in which individuals have fixed preferences over groups. It also makes it possible to derive new welfare implications.

In our model, players belong to different groups that differ in their mental models, i.e., perspectives, interpretations, narratives, and worldviews [25]. Mental models in turn shape unwritten rules or customs that prescribe the appropriate course of action. Importantly, prescriptions are not universal; rather, they are

*We thank Sandeep Baliga, Vincent Crawford, Georgy Egorov, Tim Feddersen, Karla Hoff, Matthew Jackson, Rachel Kranton, Pooya R. Ravari, Yuval Salant, Paola Sapienza, Rajiv Sethi, Eran Shmaya, Andy Skrzypacz, Jakub Steiner, Jeroen Swinkels, and numerous seminar audiences and conference participants for helpful comments and stimulating discussions. We are grateful to the Associate Editor and two referees for insightful comments that have significantly improved the paper. Alex Limonov provided research assistance.

*Corresponding author
Email addresses: willemien.kets@economics.ox.ac.uk (Willemien Kets), sandroni@kellogg.northwestern.edu (Alvaro Sandroni)
[1]Department of Economics, University of Oxford, and External Faculty, Santa Fe Institute
[2]Managerial Economics and Decision Science, Kellogg School of Management, Northwestern University.

situational. For example, offering to share food may be appropriate in some settings (e.g., family dinners) and not in others (e.g., business dinners), while in others it is unclear (e.g., social outings with colleagues). Individuals who share the same mental models tend to agree on what is the salient prescription, and this facilitates social interactions [28]. This may give players an incentive to associate with members of their own group.

These ideas have a long history in economics [56] and philosophy [33], and have recently received some experimental support [59, 37]. However, they have been difficult to model formally as the standard approach in economics leaves no room for mental models. Following our earlier work [38], we therefore depart from the standard approach by explicitly modeling how mental models shape players' reasoning about other players. We do so by building on research in psychology on theory of mind. "Theory of mind" is the cognitive capacity to attribute mental states to other people. An important component of theory of mind is introspection: to take another person's perspective, players observe their own mental state and project it onto the other. Our central assumption is that a player's own mental state is more informative of the mental states of members of his own group than of those of other groups.[3] As we show, this makes it easier to anticipate the actions of members of one's own group. Accordingly, players face less strategic uncertainty when interacting with their own group.

A first observation is that this reduction in strategic uncertainty can be beneficial when players' primary motivation is to coordinate their actions. Accordingly, players may have an incentive to seek out members of their own group. This leads us to consider an extended game where players can seek out members of their own group by choosing the same *project* (e.g., a hobby, profession, or neighborhood) before playing a coordination game with players who have chosen the same project. Players have a dual objective: On the one hand, they have an intrinsic preference over projects. On the other hand, they have an incentive to choose the same project as other members of their group in order to reduce strategic uncertainty.

We show that players may choose to seek out similar others even if that means choosing a project that they do not intrinsically prefer. The resulting level of homophily can be high. In fact, the level of homophily always exceeds that based on intrinsic preferences. In addition, the level of homophily is higher when players benefit more from reducing strategic uncertainty (i.e., coordination payoffs are high) and when interacting with the own group has a greater impact on strategic uncertainty (i.e., when a player's mental state is relatively more informative of that of members of his own group). While these predictions are intuitive, they are difficult to obtain with the standard game-theoretic framework, as the game has many equilibria and standard refinements have no bite. Our novel behavioral approach resolves this problem in a natural way: the introspective process by which players take others' perspective "anchors" players' beliefs, and this leads to a unique prediction.

Our model can explain homophily based on values, attitudes, and beliefs, which is a primary determinant of social interactions [48], and it is consistent with experimental evidence in social psychology that attitude, belief, and value similarity lead to attraction and interaction [34]. This may help explain why we might observe segregation along apparently payoff-irrelevant dimensions such as religion: While there may not be any direct payoff benefits associated with interacting with the same group, there may be strategic advantages, in the form of a reduction of strategic uncertainty.

At a deeper level, the model can shed light on why homophily may be context-dependent. For example, a well-documented phenomenon is that homophily on the basis of race is reduced substantially when individuals

[3]For evidence from psychology, see, e.g., [26, 31, 51, 60]. This assumption rules out cases where players from different groups have highly correlated but different mental states. See Sections 2 and 5.

are similar on some other dimension such as socioeconomic status [53]. This is difficult to explain with models where individuals have an immutable dislike of certain groups, but is consistent with our model if individuals experience less strategic uncertainty if they have more factors in common. A related phenomenon is that work teams that are characterized by "faultlines," i.e., teams whose members are either similar along all dimensions or have nothing in common, experience more conflict and misunderstandings than teams with the same level of diversity but where members' characteristics are not aligned [44]. This is again consistent with our model if some overlap in characteristics reduces strategic uncertainty. Our model may also help understand why individuals sometimes have a tendency to identify with groups that strongly distinguish themselves in their values and practices, even if these distinctions are valued negatively [7, 27]. In our model, distinctive practices may reduce strategic uncertainty, making it attractive to join the group even when the practices themselves are dysfunctional.

Homophily has often been linked to segregation and inequality [36] and could thus have adverse welfare consequences. This leads us to consider the question whether reducing homophily can improve social welfare. Rather than focusing on particular public policies, we take a general approach and compare the socially optimal level of homophily to the equilibrium level. Somewhat surprisingly, policies that aim to reduce homophily cannot improve social welfare in the benchmark model. This remains largely true when there are skill or informational complementarities across groups so that players have an economic incentive to interact with the other group. In that case, reducing homophily may improve welfare, but only by an arbitrarily small amount. So, a basic welfare analysis suggests that policy interventions that reduce homophily can have at best a minimal positive impact on social welfare, and can even be detrimental.

While our model is admittedly stylized, these findings raise the question how we can understand the prevalence of policies and programs that aim to reduce homophily.[4] One obvious answer is that it can be desirable to reduce homophily for noneconomic reasons. While noneconomic rationales can certainly be important, there could also be an economic case for policies to reduce homophily if there are any nonstandard frictions that cause welfare loss. One potential source of friction in our model is that players face strategic uncertainty. This friction is absent from the standard framework since standard models assume away strategic uncertainty by assuming that players coordinate on one of the Nash equilibria. However, if players have an imperfect understanding of others' mental models, as in our model, then there is a potential for miscoordination (i.e., mismatched choices). Moreover, since players find it easier to put themselves into the shoes of players who are similar to themselves, the scope of miscoordination is greater when players belong to different groups.

We find that there is indeed a "wedge" between the marginal benefit of interacting with the own group when there is strategic uncertainty and the hypothetical marginal benefit in the absence of strategic uncertainty. This wedge distorts players' incentives, thus creating a potential inefficiency. Commonly-used policies to create an inclusive culture may reduce the wedge.[5] Such policies can, but need not, be welfare-improving, depending on the precise mechanism by which they operate. Policies that make the groups' cultures more similar (i.e., reduce the cultural distance between groups) have an unambiguously positive impact on social welfare as they reduce the wedge without increasing the scope for miscoordination. On the other hand, policies that keep the cultures intact but stimulate cultural assimilation (i.e., sensitivity to other groups' culture) have ambiguous

[4] For example, while incoming students are willing to exert costly effort to interact mostly with people from the same background, some universities choose to randomly assign freshmen to dorms with an eye towards facilitating interactions between students of different backgrounds [12]; see Boisjoly et al. [14] for an analysis of the effects of such policies.

[5] For example, following a merger or acquisition, companies often use "social controls" (e.g., organizing introduction programs, cross-visits, retreats, celebrations and other socialization rituals) to create a common culture [43].

effects on social welfare: they reduce the wedge, but may lead to more miscoordination if the partial loss of one's own culture makes it harder to coordinate with the own group. Which mechanism – reduction in cultural distance or cultural assimilation – prevails depends on the precise setting [13], making it difficult to predict a priori whether policies to create an inclusive culture will improve social welfare in a given setting.

These results illustrate the value of our approach: without formalizing the intuition that homophily might stem from a desire to reduce strategic uncertainty, it would be difficult to evaluate the welfare impacts of policies that directly reduce homophily or that aim to create an inclusive culture. Moreover, our findings illustrate the importance of distinguishing between economic incentives and the sociocultural environment. In our model, players trade off the benefits of reducing strategic uncertainty with the cost of doing so. For a given sociocultural environment, individual incentives and social welfare are essentially aligned, implying that policies that directly target the level of homophily can have a small positive impact at best. Policies that target the sociocultural environment (e.g., by stimulating cultural assimilation) can reduce some distortions but possibly at the expense of increasing the risk of miscoordination for at least some groups. This suggest that policies to reduce homophily are not necessarily welfare-improving even when segregation is costly.

This paper is organized as follows. We present our basic model in Section 2. Section 3 characterizes the level of homophily in the benchmark model and presents the comparative statics. Section 4 presents the results on social welfare. Section 5 compares our approach to alternative models of homophily and discusses the robustness of our approach. Section 6 discusses the broader related literature and Section 7 concludes. All proofs can be found in the appendices.

2. Coordination and introspection

There are two groups, A and B, each consisting of a unit mass of players. Members of these groups are called A-players and B-players, respectively. Group membership is unobservable.[6]

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

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

, $v > 0$.

Payoffs are commonly known. The game has two strict Nash equilibria: one in which both players choose s^1, and one in which both players choose s^2. Thus, players cannot deduce from the payoffs alone how others will behave. So, even though there is no payoff uncertainty, there is significant *strategic uncertainty*: players do not know what the opponent will do.

A standard approach in game theory is to resolve this strategic uncertainty by selecting a Nash equilibrium. Under this approach, there is no scope for homophily unless players somehow have a preference for interacting with their own group. To capture that players may want to interact with their own group because they face less strategic uncertainty when they interact with players who share the same mental model (e.g., perspective, interpretations, categories, worldviews), we therefore depart from standard game theory and instead use the concept of introspective equilibrium, introduced by Kets and Sandroni [38]. Introspective equilibrium is defined

[6] This assumption seems suitable for homophily based on values, attitudes, and beliefs [48]. If group membership is (imperfectly) observable, homophily can be even more pronounced. In effect, it is then easier for players to segregate.

by explicitly modeling players' reasoning process and how this is shaped by their mental models. The starting point is the observation of Thomas Schelling [56, p. 96], that, when facing strategic uncertainty, "[a player's] objective is to make contact with the other player through some imaginative process of introspection." To reach such a "meeting of the minds," players can use their theory of mind. Theory of mind is a central concept in psychology. It refers to the cognitive ability to take another person's perspective. This involves introspection: depending on their mental model, players may have an impulse to take a certain action (e.g., because the action is salient to them). Before taking an action, however, players introspect and form a belief about the other player's impulse. Players then reason about others' impulses using a naive understanding of psychology, which may lead them to adjust their belief [6].

To model this, we follow Kets and Sandroni [38] and assume that each player j receives an *impulse* $i_j = 1, 2$. Impulses are payoff-irrelevant, privately observed signals and are drawn from a common prior (specified below). If a player's impulse equals 1, then his initial impulse is to take action s^1. Likewise, if a player's impulse is 2, then his initial impulse is to choose action s^2. A player's first instinct is to follow his initial impulse, without any strategic considerations. Thus, the instinctive reaction of a player with impulse $i_j = 1$ is to play action s^1. More generally, the level-0 strategy σ_j^0 for player j is defined by $\sigma_j(i_j) = s^{i_j}$ for impulse $i_j = 1, 2$. Through introspection, a player realizes that the other player likewise follows his impulse. By observing his own impulse, a player can form a belief about his opponent's impulse and formulate a best response against the belief that the opponent follows her impulse, i.e., to the opponent's level-0 strategy. This defines the player's level-1 strategy σ_j^1. In general, at level $k > 1$, a player formulates a best response against his opponent's level-$(k-1)$ strategy. This, in turn, defines his level-k strategy σ_j^k. This defines a reasoning process with infinitely many levels. The levels do not represent actual choices; they are merely constructs in a player's mind. The limit of this process (if it exists) defines an *introspective equilibrium*, i.e., a strategy profile $\sigma = (\sigma_j)_j$ is an introspective equilibrium if $\sigma_j = \lim_{k\to\infty} \sigma_j^k$.

People's impulses are influenced by their mental model, so that people who share the same mental model are likely to have the same instinctive response. On the other hand, people with different mental models may respond very differently. In many settings of interest, players have an imperfect understanding of the other groups' mental models. In the foodsharing example in the introduction, a new hire may be inclined to share his appetizer with his new colleagues but may be uncertain about their reaction (even if he had no difficulties anticipating the reactions of his colleagues at his old workplace). As another example, when communicating with members of other groups, a person may be uncertain about how to interpret their communication style or nonverbal cues. This can lead to mishaps and misunderstandings, sometimes with important economic consequences. For example, mergers and acquisitions often fail to live up to expectations due to differences in communication styles [16]; and there are more misunderstandings between doctors and patients if they are of a different race [5], gender [30], or ethnic background [17], with adverse effects on morbidity and mortality rates.

To model that players find it easier to predict the instinctive reactions of members of their own group, we again follow Kets and Sandroni [38] by assuming that impulses are more strongly correlated within groups than across groups. Each group $g = A, B$ is characterized by a state $\theta_g = 1, 2$. A priori, θ_g is equally likely to be 1 or 2 for each group g. Conditional on the $\theta_g = m$, each member of g has an impulse to play action s^m with probability $q \in (\frac{1}{2}, 1)$, independently across players. Conditional on having impulse $i_j = m$, a player j assigns probability

$$Q_{in} := q^2 + (1-q)^2$$

to a member of his group having the same impulse. We thus refer to Q_{in} as the *within-group similarity index*.

The within-group similarity index lies strictly between $\frac{1}{2}$ and 1. If the within-group similarity index is close to 1, then members of the same group are likely to have the same impulse. If it is close to $\frac{1}{2}$, then the impulses of members of a given group are almost independent.

The states θ_A, θ_B are correlated but only imperfectly so: their joint distribution is given by:

	$\theta_B = 1$	$\theta_B = 2$
$\theta_A = 1$	$\frac{1}{4} \cdot (1+\eta)$	$\frac{1}{4} \cdot (1-\eta)$
$\theta_A = 2$	$\frac{1}{4} \cdot (1-\eta)$	$\frac{1}{4} \cdot (1+\eta)$

where $\eta \in [0,1)$. Then, $\delta := 1 - \eta$ can be viewed as the cultural distance between groups: If δ is close to 1, then the states are almost independent; and if δ is close to 0, then the states are almost perfectly correlated. Conditional on having impulse $i_j = m$, a player assigns probability

$$Q_{out} := \delta \cdot \tfrac{1}{2} + (1-\delta) \cdot Q_{in}$$

to a member of the other group having impulse m. We refer to Q_{out} as the *cross-group similarity index*. The cross-group similarity index lies between $\frac{1}{2}$ and Q_{in} and decreases with δ. An impulse may thus contain some information of the impulses of members of the other group (i.e., $Q_{out} \geq \frac{1}{2}$), but it is more informative of the impulses of members of the own group ($Q_{out} < Q_{in}$), and this difference is more pronounced for groups that are not culturally close.

As shown by Kets and Sandroni [38], every introspective equilibrium is a correlated equilibrium, so that, by the epistemic characterization of Aumann [8], behavior in an introspective equilibrium is always consistent with common knowledge of rationality.[7] So, while introspective equilibrium is based on ideas from psychology and assumes that players' initial reaction is nonstrategic, it does not presume that players are irrational.

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

Proposition 2.1. [Introspective Equilibrium Pure Coordination Game] *The pure coordination game has a unique introspective equilibrium. In this equilibrium, each player follows his initial impulse.*

In this case, the introspective process thus delivers a simple answer: it is optimal to act on instinct. The intuition is straightforward: Suppose a player has an impulse to choose action $s = s^1, s^2$. Then, through introspection, he realizes that the other player is also likely to have an impulse to choose s (as $Q_{in} > \frac{1}{2}, Q_{out} \geq \frac{1}{2}$). So, at level 1, it is optimal for him to follow his impulse. But of course the same holds true for the other player. Given this, it is optimal for both players to follow their impulse at any level k, and the result follows. Hence, the initial appeal of following one's impulse is reinforced at higher levels, through introspection: as a player realizes that the other player follows her impulse, it is optimal for him to do so as well; this, in turn, makes it optimal for the other player to follow her impulse.

Proposition 2.1 demonstrates that introspection anchored by impulses makes it possible for players to coordinate by breaking the symmetry between actions. Intuitively, impulses act as a coordinating signal in this case, with the similarity indices (i.e., Q_{in}, Q_{out}) measuring the probability that players have received the same signal. However, coordination is not perfect: since impulses are imperfectly correlated, the coordination rate lies strictly between 50% and 100%. So, if players are introspective, then the mixed-Nash prediction (50%) for the coordination rate is overly pessimistic, while the pure-Nash prediction (100%) is too optimistic.

[7] However, unlike in correlated equilibrium, players need not follow their impulse in introspective equilibrium; see, e.g., Section 3.

Moreover, the coordination rate is higher when players are likely to have similar impulses (i.e., Q_{in}, Q_{out} high). These predictions are consistent with experimental evidence, which shows that behavior in coordination games is generally not consistent with Nash equilibrium, and the coordination rate is higher when one of the alternatives is highly salient [11, 49].

In addition to providing intuitive predictions, introspective equilibrium has the attractive feature that it yields a unique prediction even though the coordination game has many (correlated or Nash) equilibria and despite the fact that standard refinements have no bite in this environment. The uniqueness of introspective equilibrium will be critical for deriving unambiguous comparative static results and welfare implications in Section 3 and 4, respectively.

3. Homophily

We next turn to the question whether players' desire to reduce strategic uncertainty can explain homophily. As a first step, notice that, by Proposition 2.1, players' expected payoff in the unique introspective equilibrium is:

$$\left[\hat{p} \cdot Q_{in} + (1 - \hat{p}) \cdot Q_{out}\right] \cdot v, \tag{3.1}$$

where $\hat{p}$ is the probability that a player is matched with a member of his own group. Hence, the *marginal benefit of interacting with the own group* is given by

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

Since $Q_{in} > Q_{out}$, the marginal benefit of interacting with the own group is strictly positive. Proposition 2.1 thus has the following corollary:

Corollary 3.1. *A player's expected payoff strictly increases with the probability $\hat{p}$ of being matched with a player from the own group.*

Intuitively, players are more likely to coordinate with members of their own group, consistent with experimental evidence [59]. There are two effects. First, a player's own impulse is more informative of the impulse of a member of his own group than that of a member of the other group. Second, members of the same group are likely to have the same impulse. In the present environment, these two effects ensure that players benefit from interacting with the same group.[8]

Corollary 3.1 implies that players have an incentive to associate primarily with members of their own group, that is, to be homophilous. Hence, the matching probability $\hat{p}$ will generally be endogenous. One way people can seek out similar others is by choosing a common location or activity. We thus consider 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 with a player who has chosen the same project (uniformly at random). Once matched, players play the coordination game described in Section 2.

Each player has an intrinsic value for the projects. Players in group A have a slight intrinsic preference (on average) for project a while players in B have a slight preference for project b. More precisely, 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

[8] This need not be true for other games. For example, in games where players want to surprise the opponent, having the same impulse may not be beneficial. Moreover, the marginal benefit of interacting with the own group can be negative if there are economic benefits associated with interacting with the other group, as in Section 4.2. And being able to predict the opponent's impulse may actually be harmful even in games with a coordination motive where groups are identical in terms of payoffs if there is scope for inefficient lock-in [38].

drawn uniformly at random from $[0, 1-2\varepsilon]$, for some small $\varepsilon > 0$; the analogous statement holds for players in group B with the roles of projects a and b reversed. Values are drawn independently (across players, projects, and groups). Under these assumptions, a proportion $\frac{1}{2}+\varepsilon$ of A-players intrinsically prefer project a, and a proportion $\frac{1}{2}+\varepsilon$ of B-players intrinsically prefers project b (see Appendix A.1 for details). Thus, project a is the *group-preferred project* for group A, and project b is the group-preferred project for group B.

A player's payoff is the sum of the intrinsic value of his chosen project and his (expected) payoff in the unique introspective equilibrium of the coordination game. By (3.1), the expected payoff of an A-player with project a is

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

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

To choose their project, players follow the same introspective process as before, taking into account their payoffs in the coordination game in the second stage. At level 0, players select the project they intrinsically prefer.[9] This defines a level-0 strategy, as before. At level $k > 0$, players formulate a best response to the level-$(k-1)$ strategy of other players: a player chooses project a if and only if the expected payoff from project a is at least as high as from b, given the player's intrinsic preferences and the level-$(k-1)$ strategies of other players. Let p_{k-1}^a be the probability that an A-player at project a is matched with an A-player if other players follow the level-$(k-1)$ strategy; likewise, let p_{k-1}^b be the probability that a B-player at project b is matched with a B-player under the level-$(k-1)$ strategy profile. Then, at level k, choosing project a is a best response for an A-player if and only if

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

and analogously for other combinations of groups and projects. This defines the level-k strategy. Again, the strategies at different levels do not represent actual decisions; they are merely constructs in the players' minds. The limit of this process is again an introspective equilibrium. The limiting behavior of p_k^a and p_k^b is well-defined:

Lemma 3.2. [Convergence of Introspective Process] *The sequence $p_0^\pi, p_1^\pi, \ldots$ has a unique limit p^π for each project $\pi = a, b$. Moreover, the limits for the two projects coincide: $p^a = p^b$.*

Figure 1 illustrates the convergence of the introspective process. We write $p := p^a = p^b$ for the limiting probability. That is, p is the probability in introspective equilibrium that a player with the group-preferred project is matched with a player from the same group. Equivalently, it is the proportion of players at project a (resp. project b) who belong to group A (resp. group B) in introspective equilibrium.

The *level of homophily* is the difference $h := p - \frac{1}{2}$ between the probability p that a player with the group-preferred project is matched with player from the same group in the introspective equilibrium and the probability that he is matched with a player from the same group independent of project choice (i.e., uniformly at random). The level of homophily reflects the degree of segregation: If the level of homophily is close to $\frac{1}{2}$, then there is nearly complete segregation in the sense that the proportion of players interacting primarily with their own group is close to 1. Conversely, if the level of homophily is close to 0, then there is almost full integration in the sense that players are about equally likely to interact with both groups.[10]

[9] Alternatively, one could assume that players have an initial impulse to choose a certain project (where impulses are again more strongly correlated within groups than across groups). Our results go through in this alternative model.

[10] This is also reflected in the ex ante probability that players interact with members of their own group, which is given by $p^2 + (1-p)^2$ and which increases in p for $p \geq \frac{1}{2}$.

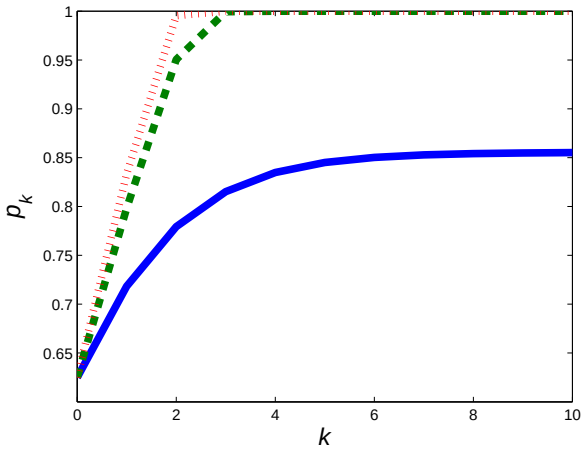

Figure 1: The probability p_k that a player who chooses the group-preferred project is matched with a member of his own group under the level-k strategy profile as a function of k, for $\beta = 0.4$ (solid line), $\beta = 0.8$ (dashed line), and $\beta = 1$ (dotted line), for $\varepsilon > 0$ small.

Since there is only a slight asymmetry in preferences, the level of homophily based on intrinsic preferences is minimal: $h^0 := \varepsilon$. The next result shows that the equilibrium level of homophily can nevertheless be high:

Proposition 3.3. [Homophily: Equilibrium] *There is a unique introspective equilibrium of the extended game. In the unique equilibrium, players follow their impulse in the coordination game, and players' project choices lead to complete segregation ($h = \frac{1}{2}$) if and only if*

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

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

$$h = \frac{(1-2\varepsilon)}{4\beta^2} \cdot \left[2\beta - 1 + \sqrt{\frac{4\beta^2}{1-2\varepsilon} - 4\beta + 1}\right]. \tag{3.2}$$

In any case, the equilibrium level of homophily exceeds the initial level of homophily (i.e., $h > h^0$).

Proposition 3.3 shows that there can be substantial homophily in the unique introspective equilibrium. In that case, most players choose the group-preferred project even if they have a strong intrinsic preference for the other project. Interactions may thus be homophilous even when players have no direct preference for interacting with their own group. Indeed, homophily is not the result of any payoff-relevant differences between groups: groups are almost identical (i.e., ε is arbitrarily small), and if homophily were based solely on intrinsic preferences, then homophily would be negligible (i.e., $h = h^0 = \varepsilon$). Instead, homophily is the result of strategic considerations. Strategic considerations *always* produce more homophily than would follow from differences in intrinsic preferences over projects (i.e., $h > h^0$), independent of the distribution of impulses and the specific assumptions on intrinsic preferences. In this sense, introspection and players' desire to reduce strategic uncertainty are root causes of homophily.

Our model thus helps explain value homophily. Value homophily is a central concept from sociology. It refers to homophily based on values, attitudes, and beliefs. Value homophily may drive homophily among other dimensions (e.g., demographic or socioeconomic) and is thus a primary determinant of social interactions [48]. The high level of homophily predicted by Proposition 3.3 is consistent with experimental evidence in social psychology that similarity in beliefs leads to attraction and interaction [34]. Our model thus provides microfoundations for this central form of homophily.

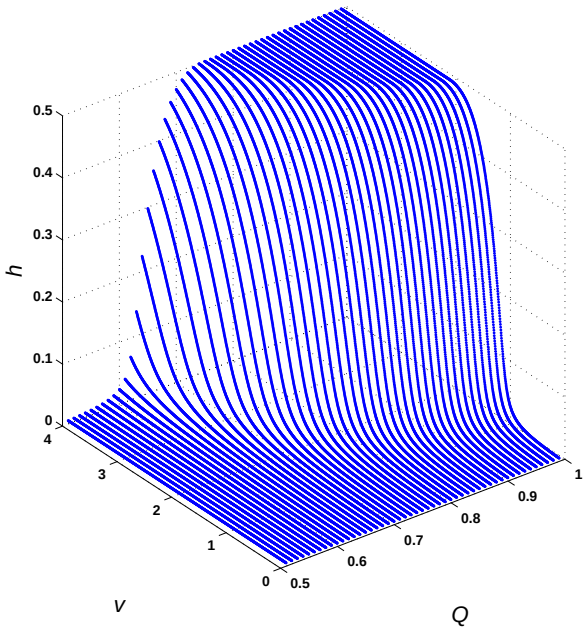

Figure 2: The equilibrium level of homophily h as a function of the coordination payoff v and the within-group similarity index Q_{in}.

Notice that, unlike in the coordination game in Section 2, players do not necessarily follow their impulse when choosing a project. Instead, introspection leads them to reconsider their initial response. At level 1, player realize that the likelihood of interacting with members of their own group is higher if they choose the group-preferred project. Players with a slight preference for the other project may thus decide to choose the group-preferred project at level 1. This further increases the likelihood of interacting with the own group when choosing the group-preferred project. Accordingly, the proportion of players choosing the group-preferred project at level 2 may be even higher. So, the attractiveness of the group-preferred project is reinforced throughout the entire reasoning process, as illustrated in Figure 1. As a result, the equilibrium level of homophily strictly exceeds the initial level (i.e., $h > h^0$).[11]

The comparative statics for the level of homophily follow directly from Proposition 3.3:

Corollary 3.4. [Homophily: Comparative Statics] *The level of homophily h increases with the coordination payoff v and the relative similarity in impulses within a group (i.e., $Q_{in} - Q_{out}$). The two are complements: that is, homophily is high when the coordination payoff v or the relative similarity in impulses within groups ($Q_{in} - Q_{out}$) is high (or both).*

Figure 2 shows the level of homophily as a function of the coordination payoff v and the within-group similarity index Q_{in} (for fixed Q_{out}). Regardless of the similarity in impulses within and across groups (i.e., Q_{in}, Q_{out}), the level of homophily increases with the economic incentives to coordinate. These comparative statics results deliver clear and testable predictions: there is a positive correlation between coordination payoffs and homophily, regardless of the exact distribution of impulses (provided that $Q_{in} > Q_{out}$).

Corollary 3.4 also demonstrates that groups that have more similar impulses (i.e., $Q_{in} - Q_{out}$ high) are more homophilous. This suggests that a similarities can be reinforcing: if members of the same group are more similar than members of different groups (i.e., $Q_{in} - Q_{out}$ high), then members of the same group tend to choose the same project; this, in turn, may lead to more shared experiences and mutual influence, leading

[11] Some other models of segregation also feature self-reinforcing dynamics. The connection to our model is strenuous at best, both in terms of modeling assumptions and results. For example, in the tipping-point model of Schelling [57], complete segregation is the only viable long-run outcome [61].

them to become even more similar.

The level of homophily also increases when groups are more distinct in terms of intrinsic preferences over projects: if a greater proportion of players has an intrinsic preference for the group-preferred project (i.e., ε increases), then there is more scope for segregation (i.e., the range of β for which there is full segregation expands); moreover, when segregation is not complete (i.e., the level of homophily is given by (3.2)), the level of homophily increases with ε. However, a distinctive feature of our approach is that even when groups are almost identical (i.e., ε is arbitrarily small), there can be substantial homophily if the benefits of reducing strategic uncertainty are large (i.e., β sufficiently large).

The introspective process plays a critical role in deriving these predictions. In particular, they are difficult to derive in standard equilibrium models where players have a direct preference for interacting with their own group. For example, consider the following model: players have an intrinsic preference over projects, and, in addition, their payoff from a project increases if more members of their group choose the project. As any game with coordination motives or network effects, this model has multiple Nash equilibria. Consequently, the comparative statics are difficult to analyze: the set of (correlated or Nash) equilibria of the game may change when payoffs are varied. In particular, depending on the exact specification of the model, the effects of parameter changes need not be monotone (see Appendix B for details). By contrast, the introspective process selects a unique equilibrium, making it possible to derive testable implications.

The introspective process thus allows us to formalize the idea that players have an incentive to interact with members of their own group because they know what to expect of them. The comparative statics confirm this intuition by showing that the level of homophily varies with both economic factors (i.e., v and ε) and the sociocultural environment (i.e., Q_{in}, Q_{out}) in an intuitive way. The model thus sheds light on the minimal assumptions on primitives under which a desire to reduce strategic uncertainty leads players to seek out others who are similar to themselves.

In addition, the model can shed light on a number of empirical regularities. In particular, it suggest why homophily can be situational, with individuals displaying only a weak preference for their own demographic group if individuals of other demographic groups are similar on socioeconomic dimensions [53]. This can be explained using a slightly richer model where identity has multiple dimensions (e.g., socioeconomic, age, race), and the correlation in impulses among two individuals increases with the number of dimensions they have in common. For example, suppose that there are two dimensions that can each take on one of two values so that there are four "types" of players, labeled $t_{11}, t_{12}, t_{21}, t_{22}$ (where type t_{RS} belongs to "subtype" R on the first dimension and to subtype S on the second dimension). Suppose also that players are more likely to have the same impulse if they have more dimensions in common.[12] Then, players benefit most from interacting with players who are similar on both dimensions (i.e., type t_{RS} receives the highest expected payoff from interacting with type t_{RS}) but may benefit almost as much from interacting with players who are similar on only one dimension, depending on the correlation in impulses along this dimension and on how central this dimension is to a player's identity. For example, the expected payoff of type t_{RS} from interacting with type $t_{R'S}$ (for $R' \neq R$) may be close to the expected payoff of interacting with type t_{RS} if the second dimension is especially salient.

[12] For example, suppose that there are four states, labeled $\theta_{R^1}, \theta_{R^2}, \theta^{S^1}, \theta^{S^2}$, which can be either 1 or 2, and that are (imperfectly) correlated, as before. Each player's impulse is shaped by the first dimension with probability $p_d \in [0,1]$; if a player has type t_{RS} and his impulse is shaped by the first dimension, then, conditional on $\theta_R = m$, he has an impulse to choose action s_m with probability $q > \frac{1}{2}$; and likewise if his impulse is shaped by the second dimension. The parameter p_d can be interpreted as measuring the importance or salience of the first dimension for a player's identity. It is then straightforward to derive the marginal benefit β of interacting with the own type for this environment; all our results then go through with this more general definition.

This extension of the model can also explain the important phenomenon that misunderstandings and conflict are more prevalent when there are "faultlines" in a group (e.g., a work team), where faultines occur if most players are either similar along both dimensions or have nothing in common [44]. In the extension of the model described above, faultlines can be captured by assuming that most players are either of type t_{11} or of type t_{22} (but few of type t_{12} or t_{21}). Then, our model predicts that players have a strong incentive to segregate; and if they cannot segregate (e.g., because they belong to the same organizational department), their expected payoff in the coordination game are lower than if characteristics were not aligned (i.e., if a sizeable proportion of players are of type t_{12} or of type t_{21}), capturing the idea that faultlines lead to more mishaps and misunderstandings.

Another empirical phenomenon that the model can readily explain is that individuals may identify with groups that have distinctive practices such as countercultures or disaffected groups even if these practices are negatively valued [7]. This can be understood in our model if distinctive practices reduce strategic uncertainty. For example, consider an extension of the model where players who choose project a play a coordination game with payoff v, as before, but players who choose project b receive only $v' \in (0, v)$ if they coordinate (and 0 otherwise). In this model, members of a certain group may overwhelmingly choose project b if it is easier for them to anticipate players instinctive responses in the low-payoff game. That is, suppose that the probability that members of the same group have the same impulse in project a and b is

$$\begin{aligned} Q_{in}^a &= q_a^2 + (1 - q_a)^2; \\ Q_{in}^b &= q_b^2 + (1 - q_b)^2; \end{aligned}$$

respectively, where $q_b > q_a$. Then, the expected payoff of playing the low-payoff coordination game, $Q_{in}^b \cdot v'$, may exceed the expected payoff in the high-payoff coordination game, $Q_{in}^a \cdot v$. If that is the case, arbitrarily small differences between groups (e.g., in intrinsic preferences over projects or in impulses) may be amplified and may lead one group to coordinate on activities with inherently low payoffs. If coordination problems diminish over time (i.e., $Q_{in}^a, Q_{in}^b \to 1$), then this group may be worse off in the long run (i.e., $v' < v$) and may be locked into a low-payoff state if switching projects is costly. These insights are difficult to derive using model in which players have a direct preference for interacting with their own group. For example, if players have a sufficiently strong direct preference for interacting with their own group, players are always better off under segregation.

4. Social welfare

While individual players may benefit from interacting with their own group, a policy-maker might be concerned that the resulting segregation can be socially inefficient. This motivates us to characterize the socially optimal level of homophily and to compare it with the equilibrium level. We first consider the benchmark model, and then consider an extension of the model where players may benefit from interacting with the other group. We conclude by considering nonecomic policies that aim to change the sociocultural environment so as to influence the strategic uncertainty that players face when interacting with different groups.

4.1. Benchmark model

We characterize the project allocations that maximize social welfare (i.e., the sum of coordination payoffs and project values). We compare the socially optimal level of homophily to the equilibrium level. The analysis is greatly simplified by the uniqueness of the introspective equilibrium.

We start with defining the socially optimal level of homophily. Clearly, for any given level of homophily, it is socially optimal to assign players with the highest intrinsic preference for a given project to that project. Therefore, any level of homophily defines to a unique cutoff value for the intrinsic preferences such that players whose (relative) preference for the group-preferred project exceeds the cutoff are assigned to that project. Given this, the social optimum is completely characterized by the level of homophily. Thus, the *socially optimal level of homophily* h^* is the level of homophily that maximizes social welfare $W(h)$, given by

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

where $C(h)$ is the total coordination payoff and $\Pi(h)$ is the total value that players assign to projects (under the cutoff corresponding to h). The total value $\Pi(h)$ that players derive from projects is characterized in Appendix A.2. Clearly, the total project value $\Pi(h)$ is maximized if all players choose the project that they intrinsically prefer (i.e., $h = h^0$) and is a decreasing function of the deviation $|h - h^0|$. The total coordination payoff is characterized by the following lemma:

Lemma 4.1. *If the level of homophily is h, then the total coordination payoff is*

$$C(h) = 2v \cdot [Q_{in} \cdot \left((\tfrac{1}{2} + h)^2 + (\tfrac{1}{2} - h)^2\right) + Q_{out} \cdot (\tfrac{1}{2} + h) \cdot (\tfrac{1}{2} - h)].$$

If players' primary motive for interacting with similar others stems from their desire to reduce strategic uncertainty, then high levels of homophily can be socially optimal, as the next result shows:

Proposition 4.2. [Social Welfare] *Full segregation is socially optimal (i.e., $h^* = \frac{1}{2}$) if and only if*

$$\beta \geq \tfrac{1}{2} - \varepsilon. \tag{4.1}$$

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

$$h^* = \frac{(1-2\varepsilon)}{4\beta^2} \cdot \left[\beta - \frac{1}{4} + \sqrt{\frac{4\beta^2}{1-2\varepsilon} - \tfrac{1}{2}\beta + \tfrac{1}{16}}\right]. \tag{4.2}$$

In all cases, the fraction of players choosing the group-preferred project exceeds the initial level (i.e., $h^* > h^0$).

Proposition 4.2 characterizes the conditions under which homophily can be socially optimal. There is a tradeoff. When the level of homophily is high, players can accurately predict their opponent's impulse in the coordination game. This minimizes strategic uncertainty and maximizes players' payoff in the coordination game. However, segregation requires that some players choose a project that they do not intrinsically prefer. The social optimum trades off these two factors. If the coordination payoff v or the relative similarity in impulses within groups (i.e., $Q_{in} - Q_{out}$) is sufficiently high (i.e., (4.1) holds), then the coordination motive dominates and full segregation is optimal. If full segregation is not optimal, then the socially optimal level of homophily is pinned down by the relative similarity in impulses within and across groups (i.e., Q_{in}, Q_{out}) and the economic benefits of coordination (i.e., v). The comparative statics for the socially optimal level of homophily are similar to those for the equilibrium level: Figure 3 shows that the optimal level of homophily increases with the within-group similarity index Q_{in} and with the coordination payoff v, and the two are complements.

We next compare the socially optimal level of homophily to the equilibrium level. This indicates whether there may be a need for intervention, and, if so, of what kind.

Corollary 4.3. *The level of homophily in the unique introspective equilibrium never exceeds the socially optimal level of homophily; and if $\beta \leq 1 - 2\varepsilon$, the equilibrium level of homophily is strictly lower than the socially optimal level of homophily.*

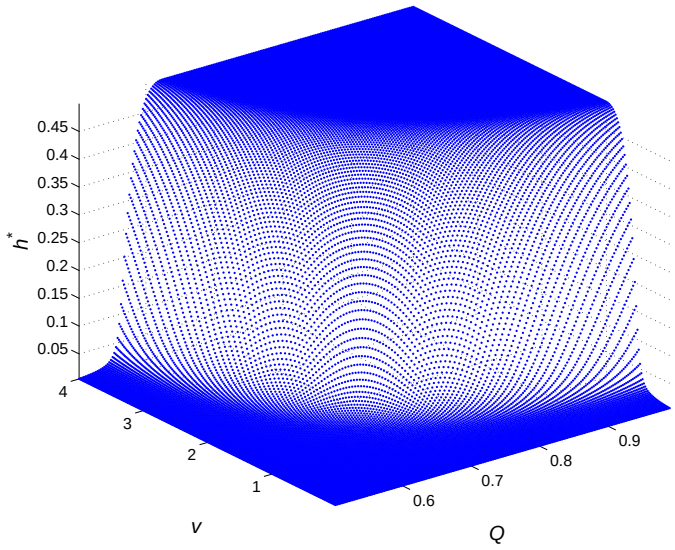

Figure 3: The socially optimal level of homophily as a function of the coordination payoff v and the within-group similarity Q_{in}.

Proposition 3.3 and Corollary 4.3 demonstrate that there are two possibilities. There may or they may not be full segregation in equilibrium. If the society is fully segregated in equilibrium, then full segregation is in fact socially optimal. Thus, an intervention cannot improve social welfare. On the other hand, if the society is not fully segregated, then it is not sufficiently segregated. So, if anything, the level of homophily is too low in equilibrium, and a welfare-improving policy *increases*, rather than decreases, the level of homophily.

The intuition behind Corollary 4.3 is as follows. There are two types of externalities that work in opposite directions. If a player switches to the group-preferred project, then this increases the expected coordination payoff for the members of his group who have chosen the group-preferred project (as it increases the probability that they interact with their own group). This is a positive externality. On the other hand, such a switch reduces the expected coordination payoff of the members of his group who have chosen the other project. This is a negative externality.[13] Since there are more players with the group-preferred project in equilibrium, the positive externality dominates the negative one at the equilibrium level of homophily (and, in fact, for any level of homophily below the social optimum). Therefore, there is too little homophily in equilibrium.

The benchmark model thus does not provide an economic rationale for policies that aim to reduce homophily. One might ask whether a similar conclusion holds if players have an economic incentive to interact with members of the other group. This is the question we turn to next.

4.2. Skill complementarities

In many situations of interest, groups have complementary skills or information so that players have a direct incentive to interact with other groups. One might conjecture that policies that reduce homophily can significantly improve welfare in this case. To address this issue, we consider a simple model where players receive a payoff $V \geq v$ if they coordinate with someone from the other group (and a payoff v if they coordinate with their own group), as in the following game:

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

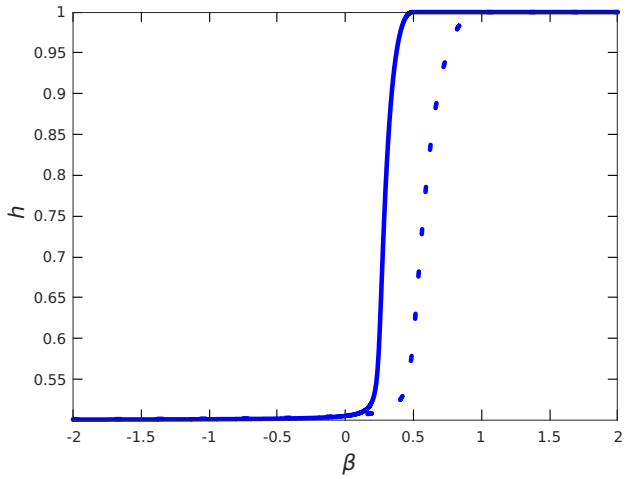

Figure 4: The socially optimal (solid line) and equilibrium (dashed) level of homophily as a function of the marginal benefit β of interacting with the own group.

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

Own group

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

Other group , $V \geq v$.

This generalizes the benchmark model, which corresponds to the case $V = v$. The case $V > v$ models the case where there are skill complementarities across groups. In this case, players always benefit from coordinating, but more so if they coordinate with a member of the other group.

Players follow the same introspective 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 the level-$(k-1)$ strategies: a player chooses project a if and only if the expected payoff from a is at least as high as from b, given the level-$(k-1)$ strategies.

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

$$\hat{p} \cdot Q_{in} \cdot v + (1 - \hat{p}) \cdot Q_{out} \cdot V.$$

The *marginal benefit of interacting with the own group* in this extended model is thus

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

The marginal benefit of interacting with the own group can be positive or negative, depending on economic incentives (i.e., the magnitude of v and V), skill complementarities (i.e., the relative magnitude of V versus v), and the similarity in impulses within and across groups (i.e., Q_{in}, Q_{out}). If skill complementarities are weak (so that $\beta > 0$), then it is easy to check that all results for the benchmark model extend. In particular, there is too little homophily in equilibrium. So, we focus here on the case where skill complementarities are strong (i.e., $\beta < 0$). The next result shows that in this case, there may be too much homophily in equilibrium; however, the deviation from the socially optimal level is minimal.

Proposition 4.4. [Social Welfare: Skill Complementarities] *If skill complementarities are sufficiently strong (i.e., $\beta < 0$), then the level of homophily in the unique introspective equilibrium is never below the socially optimal level of homophily, and may be strictly greater. However, the difference between the equilibrium level and the socially optimal level of homophily is at most ε.*

The result is illustrated by Figure 4. Figure 4 compares the socially optimal and equilibrium level of homophily. If there are no or limited skill complementarities across groups (i.e., $\beta > 0$, Proposition 4.2), then the equilibrium level of homophily is never greater than the socially optimal level, as noted earlier. In fact, if β is not too large, it can be strictly smaller. If there are strong skill complementarities across groups (i.e., $\beta < 0$, Proposition 4.4), then the equilibrium level of homophily is below the socially optimal level. However, the difference is at most ε.[14]

Proposition 4.4 thus suggests that reducing segregation can improve welfare when there are significant complementarities of skill, consistent with other arguments [e.g. 3, 32, 45, 52]. However, the difference between the socially optimal and equilibrium level of homophily is arbitrarily small in our setting unless there is significant heterogeneity in preferences. The intuition is that the externalities largely cancel each other out in our model. If there are strong skill complementarities across groups, an A-player who chooses the group-preferred project a exerts a negative externality on A-players with project a (and a positive one on B-players with project a), as well as a positive externality on A-players who choose project b (and a negative one on B-players whoc choose b), and likewise for B-players that choose project b (as $\beta < 0$). However, in the presence of skill complementarities, players have an incentive to form integrated groups, and, in equilibrium, the proportion of players experiencing a negative externality is about as large as the proportion of players experiencing a positive externality. Consequently, the two types of externalities essentially cancel out. Thus, skill complementarities provide only a weak economic rationale for policies to reduce homophily.[15]

Our model is admittedly stylized, and more complex models might yield somewhat different conclusions, depending on the specific assumptions made. However, our findings continue to hold under various alternative specifications, as we discuss in Section 6, suggesting that the main insight is robust: a basic welfare analysis does not provide a strong economic rationale for policies that reduce homophily when homophily reduces strategic uncertainty.

4.3. Sociocultural factors

In the previous section, we showed that a basic welfare analysis at best provides a weak economic rational for reducing homophily. However, basic welfare analyses – which take the sociocultural environment (i.e., Q_{in}, Q_{out}) as given – do not take into account the effect of strategic uncertainty. Notice that in the absence of any strategic uncertainty and coordination frictions, the marginal benefit of interacting with the own group is just the opportunity cost of interacting with the own group, that is,

$$\beta^* := v - V.$$

This hypothetical marginal benefit needs to be contrasted with the actual marginal benefit

$$\beta = Q_{in} \cdot v - Q_{out} \cdot V$$

when players face strategic uncertainty. Hence, strategic uncertainty drives a wedge between the opportunity cost β^* and the marginal benefit β. In particular, it could be that the opportunity cost is negative (i.e., $V > v$), while the marginal benefit is positive when strategic uncertainty is taken into account (i.e., $\beta > 0$). This is the case if players are much better at predicting the impulse of members of their own group (i.e., $Q_{in} - Q_{out}$ is large). In this case, there can be high levels of homophily (Proposition 3.3) even though the presence of skill

[14] As Figure 4 suggests, the socially optimal and equilibrium level of homophily are continuous in β.

[15] For example, in the presence of skill complementarities and significant preference heterogeneity (ε large), a planner who cares only about coordination (and not about intrinsic preferences) might wish to introduce policies to reduce homophily.

complementarities makes it optimal for players to interact with the other group if strategic uncertainty were to be eliminated.

More generally, the wedge $\beta^* - \beta$ reflects two types of inefficiencies. First, as impulses are not perfectly correlated (i.e., $Q_{in}, Q_{out} < 1$), there can be *miscoordination*: players may fail to coordinate their actions. Second, players' incentives at the project-choice stage are distorted because it is easier for players to interact with their own group (i.e., $Q_{in} > Q_{out}$). This may lead players to choose a project that they do not intrinsically prefer, which is socially costly.

This suggests that there may be room for welfare-improving interventions that target the sociocultural environment (i.e., Q_{in} and Q_{out}). However, since there are two types of inefficiencies, there may be a tradeoff. Miscoordination can be reduced, for instance, by making players more familiar with the cultural code of their own group (i.e., increasing Q_{in}). However, this might increase the distortions at the project-choice stage if players do not also become better acquainted with the cultural code of the other group (i.e., if $Q_{in} - Q_{out}$ increases).

However, some interventions may avoid this tradeoff. One possibility is to develop an inclusive culture. While this term can have many different interpretations, in the context of our framework an inclusive culture can be viewed as one where players have similar impulses regardless of which group they belong to (i.e., a small gap between Q_{in} and Q_{out} so that β is close to β^*). A proper evaluation of the welfare effects of developing an inclusive culture requires extending our model to a dynamic setting as changes in players' impulses presumably take time, making it necessary to consider intertemporal tradeoffs. Nevertheless, some useful insights can be gleaned from our model. For example, suppose some players choose a project that they do not intrinsically prefer in equilibrium (i.e., $h > h^0$). Then social welfare can be improved if players become better acquainted with the cultural code of the other group without losing familiarity with their own cultural code (i.e., $Q_{out} \uparrow Q_{in}$). This reduces the distortions in incentives at the project-choice stage (as $\beta \to \beta^*$) so that more players can choose the project that they intrinsically prefer. Moreover, the miscoordination rate is lower since players are better able to predict their opponent's impulse in the coordination game for any level of homophily. However, if the increased familiarity with the cultural code of the other group comes at the expense of knowledge of the own cultural code (i.e., Q_{in} decreases), then the welfare implications are ambiguous because we encounter the same tradeoff as before: while reducing the gap between Q_{in} and Q_{out} reduces the distortions in incentives at the project-choice stage, the increase in strategic uncertainty may lead to more miscoordination.

The possibility that an inclusive culture leads to more miscoordination depends on the mechanism by which an inclusive culture is created. One possible mechanism by which a policy can create an inclusive culture is by *reducing the cultural distance between groups* (i.e., $\delta \downarrow$): since the cross-group similarity index Q_{out} increases when the cultural distance falls while the within-group similarity index Q_{in} is independent of cultural distance, a decrease in cultural distance has an unambiguously positive effect on social welfare: it reduces the distortion at the project-choice state (by reducing the wedge between β^* and β) while increasing the coordination rate (by increasing Q_{out}).

This is not the only possibly mechanism by which an inclusive culture can be created, however. An alternative mechanism involves *cultural assimilation*, i.e., players become more sensitive to the cultural code of the other group. We can model this by assuming that with some probability, a player's impulse derives from the state of the other group. First suppose that only one group assimilates. Assuming assimilation requires costly effort (e.g., time to learn about the culture of the other group) and groups are identical in terms of effort cost, it is socially optimal if the minority assimilates. So, if group A forms the majority (at a given project)

and $\theta_A = m$, then an A-player has an impulse to choose s^m with probability q, as before. A B-player's impulse depends on how sensitive he is to the cultural code of the other group: with probability p^{ass}, conditional on $\theta_A = m$, he has an impulse to choose s^m with probability q (and an impulse to choose $s' \neq s^m$ with probability $1-q$); with probability $1-p^{ass}$, conditional on $\theta_B = n$, he has an impulse to choose s^n with probability q. The probability p^{ass} reflects the degree of cultural assimilation in the sense that it measures the degree to which the minority players are sensitive to the cultural code of the other group. In the benchmark model, there is no assimilation, so $p^{ass} = 0$. When there is some assimilation by the minority, then $p^{ass} > 0$. For a member of the majority group (viz., group A), the probability that his opponent in the coordination game has the same impulse is

$$Q_{in}^{ass,maj} = Q_{in}; \qquad Q_{out}^{ass,maj} = Q_{out} + p^{ass}(Q_{in} - Q_{out});$$

if his opponent belongs to the majority group and the minority group, respectively. For a member of the minority group, the probability that his opponent in the coordination game has the same impulse is

$$Q_{in}^{ass,min} = Q_{out} + P^{ass} \cdot (Q_{in} - Q_{out}); \qquad Q_{out}^{ass,min} = Q_{out} + p^{ass}(Q_{in} - Q_{out});$$

if his opponent belongs to the minority group and the majority group, respectively, where $P^{ass} = (p^{ass})^2 + (1 - p^{ass})^2$. So, if only the minority assimilates, then for any given level of homophily, members of the majority group always benefits from cultural assimilation since they face a lower miscoordination rate (i.e., $Q_{in}^{ass,maj}, Q_{out}^{ass,maj}$ increase with p^{ass}). On the other hand, the effect on the minority is ambiguous. If there is limited cultural assimilation (i.e., $p^{ass} < \frac{1}{2}$), then further assimilation leads to more miscoordination for minority players when they interact with other minority players (i.e., $Q_{in}^{ass,min}$ decreases with p^{ass}). The net effect on social welfare then depends on the probability that minority players interact with members of their own group. Only when the minority is already highly assimilated (i.e., $p^{ass} > \frac{1}{2}$) does further assimilation benefit the minority. This suggest that it may be welfare improving to compensate minority players for the miscoordination cost they incur at the initial stages of the assimilation process (e.g., using transfers by the majority players).

If both groups learn about each other's cultural code (i.e., assimilation is two-sided), the welfare effects are even more ambiguous. In this case, the probability that a player's opponent in the coordination game has the same impulse is

$$Q_{in}^{ass*} = Q_{out} + P^{own} \cdot (Q_{in} - Q_{out}); \qquad Q_{out}^{ass*} = Q_{in} - P^{own} \cdot (Q_{in} - Q_{out});$$

if his opponent belongs to the same and the other group, respectively, where $P^{own} = (p^{own})^2 + (1 - p^{own})^2$ and where $p^{own} \in [\frac{1}{2}, 1]$ is the probability that a player's impulse derives from his own group (so, $p^{own} = 1$ in the benchmark model).[16] In this case, there is a tradeoff: if there is more assimilation, then players find it easier to coordinate with one group but harder to coordinate with the other group. For example, if there is significant assimilation (i.e., p^{own} small), a player's within-group coordination rate (i.e., Q_{in}^{ass*}) is small but his cross-group coordination rate (i.e., Q_{out}^{ass*}) is high. The converse is true if there is only limited assimilation (i.e., p^{own} close to 1). Again, the net effect on social welfare depends on the level of homophily.

The above discussion suggests that whether policies to create an inclusive culture can improve social welfare depends on the precise mechanism by which they affect the sociocultural environment. Which mechanism

[16] Formally, with probability p^{own}, conditional on $\theta_g = m$, a g-player has an impulse to choose s^m with probability q (and an impulse to choose $s' \neq s^m$ with the remaining probability $1-q$); with probability $1-p^{own}$, conditional on $\theta_{g'} = n$ for $g' \neq g$, a g-player has an impulse to choose s^n with probability q. We require $p^{own} \geq \frac{1}{2}$ to rule out the case that players are more sensitive to the cultural code of the other group than to that of their own.

prevails may depend on the particulars of the broader social environment [13]. This suggests that creating an inclusive culture is not a panacea: without a better understanding of how the social environment influences which mechanism prevails, policies to create an inclusive culture may have adverse welfare consequences.

In addition, policies to create a more inclusive culture may have other weaknesses. Suppose a policy induces players to interact with members of the other group so that they become more familiar with the cultural code of the other group. Such a policy may be successful at reducing the gap between the within-group and the cross-group coordination rate so that it reduces the distortions in players' incentives at the project-choice stage. Also suppose it is implemented in such a way that it does not reduce the within-group coordination rate too much. Then, it is still not guaranteed that its welfare effects are positive. For example, it may take a great deal of time for players to learn each others' cultural code, and the costs associated with the resulting miscoordination in the short run can be high. Whether interventions can be welfare-improving then depends on the discount rate as well as on other factors, for example whether players are myopic or do not fully internalize the effect of their actions on future generations. To summarize, as in the case of skill complementarities, while our model offers some support for policies that reduce homophily, it does so only in a qualified manner.

5. Discussion

Alternative mechanisms. In our belief-based model of homophily, players have an incentive to interact with their own group because this reduces strategic uncertainty. We have contrasted the positive and normative implications of our model with those of preference-based models that assume that players have a fixed preference over groups. As discussed, our model can help explain why homophily can be context-dependent [53], why faultlines in teams can lead to more conflict [44], and why players sometimes join groups with distinctive but dysfunctional practices [7, 27]. Preference-based models can potentially explain these phenomena if the preference for interacting with the own group is sufficiently strong; however, without a theory of the determinants of these preferences, it is hard to explain why they are particularly strong in some cases, but not in others. More broadly, our approach provides testable hypotheses: homophily is high precisely when the marginal benefit of interacting with the own group is high.[17] Our model can also be used to evaluate the welfare implications of commonly-used policies to change the sociocultural environment, which requires endogenizing homophilous preferences.

Other authors have derived homophilous preferences from sources distinct from a desire to reduce strategic uncertainty. Baccara and Yariv [9] show that there can be homophily based on similarity in preferences: in their model, groups are stable only if their members have similar preferences over public goods. Our results suggest that homophily can emerge also when there is no need to invest in public goods. Pęski [55] shows that there can be segregation if players have preferences over the interactions that their friends have with other players. In Pęski's model, each player has preferences over interacting with different players: for example, player i may be friendly towards player j but hostile towards player j'; Pęski then studies the interaction patterns that can emerge under different assumptions on preferences. In both the work of Baccara and Yariv and that of Pęski, homophily is driven by preferences (over public goods and over other players, respectively), and the level of homophily is minimal when differences in preferences are small. By contrast, in our model, homophily is driven by beliefs: homophily is driven primarily by players' desire to reduce strategic

[17] This suggests a direct behavioral test of our model. Suppose subjects belong to two distinct groups that differ in which action labels are salient for their members. If subjects are matched in pairs to play a coordination game (as in Section 2), then our model predicts that subjects are willing to pay to interact with their own group if it is difficult to anticipate the salience of the action labels for members of the other group [cf. 41], but not otherwise.

uncertainty, and can be high even if groups are nearly identical in all respects (i.e., ε arbitrarily small). As a result, homophily is shaped by the interaction of economic factors (i.e., coordination payoffs and intrinsic preferences) and the sociocultural environment (i.e., cultural similarity), and homophily can be substantial even when the differences in preferences between groups is arbitrarily small (i.e., $h > h^0$). More generally, given their emphasis on differences in preferences and different focus, the models of Baccara and Yariv and Pęski do not deliver the positive and normative implications of our model highlighted above. In particular, without accounting for the effects of strategic uncertainty, it is difficult to evaluate the welfare implications of policies that influence the sociocultural environment.

The literature has also considered other reasons why individuals have a tendency to interact with similar others. The literature on peer effects emphasizes the effects of mutual influence: individuals who interact frequently become more similar over time. By contrast, in our model, being similar is a precondition for interaction, not a result thereof. Our model highlights how the two mechanisms interact: people who belong to the same group become similar on other dimensions as well, e.g., by choosing the same hobbies, professions, or clubs as other members of their group [cf. 39]. This suggests that a full understanding of peer effects requires taking into account players' incentives to interact with different groups.

Social interactions can also be shaped by opportunity [e.g., 47]. For example, tracking in schools can facilitate interactions between students of similar academic ability, and professional networks often develop between people that work at the same company or have the same expertise. Policy-makers sometimes try to reduce homophily by creating more opportunities for members of different groups to interact.[18] Our findings suggest that without a full understanding of the drivers of homophily, the welfare implications of such policies are unclear.

Group polarization. We have focused on the case where players find it difficult to anticipate the reactions of members of other groups. In other cases, members of different groups may have opposite impulses, so that there is group polarization. In this case, impulses are negatively correlated across groups: if a member of group A has an impulse to choose s, then a member of group B is likely to have an impulse to choose $s' \neq s$. Group polarization can also lead to homophily. For example, suppose that players vote on an issue, or are asked to give their opinion more generally; and suppose that members of group A have an instinctive tendency to support alternative s^A, while members of group B have an instinctive tendency to support alternative $s^B \neq s^A$. Then, if players dislike it if others support a different alternative than they do, players have an incentive to segregate into communities in which most members belong to the same group. These mechanisms are of course not mutually exclusive: in some cases, homophily might be driven by a desire to reduce strategic uncertainty, while in others, it might be driven by group polarization. In either case, the scope for homophily depends on both economic factors (i.e., payoffs) and on the sociocultural environment (i.e., impulses), underlining the importance of explicitly modeling the impact of sociocultural conditions.

Robustness. While we have derived our results for a specific environment, our results do not depend on our specific assumptions such as the exact assumptions on preferences or the distribution of impulses. In essence, the key to our results is that players can benefit from interacting with their own group. Whenever the marginal benefit of interacting with the own group is positive, there will be substantial homophily in equilibrium, regardless of the exact source of this benefit.[19] In the project choice stage, our results are robust

[18] For example, schools may create social clubs based on shared interests, universities may randomly assigning students to dorms, and companies and professional associations may organize networking events that span multiple communities; see, e.g., [14, 12].

[19] For example, similar results hold for games where players trade-off a coordination motive with their intrinsic preferences [50].

to relaxing the assumption that each project is equally attractive ex ante (e.g., both groups (intrinsically) prefer a certain project) as long as there is some asymmetry in intrinsic preferences across groups. Finally, our results go through if group membership is (imperfectly) observable or if players cannot sort by choosing projects, but instead signal their identity by choosing markers, that is, observable attributes such as tattoos or specific attire, to signal their identity and increase the chance of meeting with members of their own group (see Appendix C for details).

6. Related literature

The literature on homophily typically assumes homophilous preferences and investigates the implications for network structure and economic outcomes [e.g., 23, 15, 29], with Baccara and Yariv [9, 10] and Pęski [55] being notable exceptions, as discussed in Section 5. We propose a novel mechanism through which homophily can arise: players have an incentive to interact with similar others if that reduces strategic uncertainty.

An emerging literature in economics studies the effect of identity and culture on economic outcomes [e.g., 1, 2, 20, 24, 42]. Unlike much of the existing literature, we abstract away from any direct effects on preferences to focus on the effect on strategic uncertainty. In particular, in contrast with Akerlof and Kranton [1], "identity" is not a direct argument of a player's utility function in our formulation. Our approach allows us to formalize the idea that culture is a source of focal principles that can aid in equilibrium selection [40]. In our model, members of the same group tend to agree on focal principles, which allows them to coordinate effectively. We show that this can lead to homophily even if groups are essentially identical in all payoff-relevant aspects and players do not have a preference over groups.

The process we consider is related to the best-response process in level-k and cognitive-hierarchy models [see 21, for a survey]. There are three important differences. First, while this literature focuses on deviations from equilibrium, we use the reasoning process to select a unique equilibrium. Second, we introduce impulses that are potentially correlated. This allows us to shed light on group differences in strategic behavior and on homophily. Third, rather than explaining experimental data, our focus is on deriving testable implications that hold for any distribution of impulses (under the assumption that impulses are more strongly correlated within groups). For example, the positive implication that homophily increases with the economic benefits of interacting with the own group (Corollary 3.4) holds for any assumption on players' impulses; likewise, the normative implication that policies to reduce homophily cannot have a significant positive impact on social welfare unless they affect the sociocultural environment (Corollary 4.3 and Proposition 4.4) does not depend on the particulars of the impulse distribution. This focus allows us to draw general conclusions that are robust to specific modeling assumptions.

Modeling the introspective process allows us to select a unique outcome in a range of games. This allows us to derive clear comparative statics and new welfare implications. This is not possible using a standard equilibrium analysis (see Appendix B). Like other models that are used to explain homophily and segregation, the games we study have multiple equilibria with sometimes very different properties. Other papers have dealt with equilibrium multiplicity by focusing on the subset of equilibria that satisfy a stability property [e.g., 2]. However, these refinements have no bite in our environment, necessitating a novel approach.

Our work sheds light on experimental findings that social norms and group identity can help players coordinate effectively, as in the minimum-effort game [58, 20], communication tasks [59], the provision point mechanism [22], risky coordination games [46], and Battle of the Sexes [19, 37]. Chen and Chen [20] explain the high coordination rates on the efficient equilibrium in risky coordination games in terms of social preferences. Our model provides an alternative explanation, based on beliefs: players are better at predicting the actions of

players who belong to the same group. Our mechanism operates even if no equilibrium is superior to another in terms of payoffs.

7. Conclusions

The prevalence of homophily has long intrigued researchers across the social sciences. Rather than directly positing a preference for interacting with the own group, we derive homophilous preferences from a desire to reduce strategic uncertainty. Homophily emerges because players find it easier to predict the instinctive reactions of members of their own group. Providing microfoundations for homophilous preferences sheds light on various empirical regularities and makes it possible to derive novel welfare implications. Importantly, a high level of homophily does not, in itself, deliver an economic rationale for policies to reduce homophily. In our model, homophily is a by-product of socially valuable efforts to reduce strategic uncertainty, and homophily per se does not entail a welfare loss even if associating with similar others is costly. Indeed, even if players have an economic incentive to interact with the other group, there is only a limited economic rationale for policies that reduce homophily.

However, accounting for strategic uncertainty makes it possible to identify a new source of inefficiency: if players are uncertain about other players' actions, then there is a substantial risk of miscoordination. This risk is mitigated when players become better acquainted with others' cultural code, for example, through cultural assimilation. The net effect of such policies on social welfare can be ambiguous, however, depending, e.g., on the precise mechanism by which players become sensitive to the other group's code.

While social scientists have long acknowledged the importance of social policies that make people more familiar with other groups' codes and practices, these policies have hitherto received limited attention from economists because a formal framework to evaluate their welfare impacts was lacking. By providing a model of homophilous preferences based on beliefs, this paper opens up the possibility of studying the welfare implications of these commonly used policies. The welfare analyses in this paper, while preliminary in nature, suggest that when players face strategic uncertainty, sociocultural conditions may have a significant impact on social welfare. Further studying how the interaction of sociocultural and economic conditions affects strategic behavior and social welfare is thus a promising direction for future research.

Appendix A. Auxiliary results

Appendix A.1. Intrinsic preferences

We denote the values of an A-player j for projects a and b are denoted by $w_j^{A,a}$ and $w_j^{A,b}$, respectively; likewise, the values of a B-player for projects b and a are $w_j^{B,b}$ and $w_j^{B,a}$, respectively. As noted in the main text, the values $w_j^{A,a}$ and $w_j^{A,b}$ are drawn from the uniform distribution on $[0,1]$ and $[0,1-2\varepsilon]$, respectively. Likewise, $w_j^{B,b}$ and $w_j^{B,a}$ are uniformly distributed on $[0,1]$ and $[0,1-2\varepsilon]$. All values are drawn independently (across players, projects, and groups). So, players in group A (on average) intrinsically prefer project a (in the sense of first-order stochastic dominance) over project b; see Figure A.5. 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 trapezoidal distribution. That is, if we define $x := 1-2\varepsilon$, we

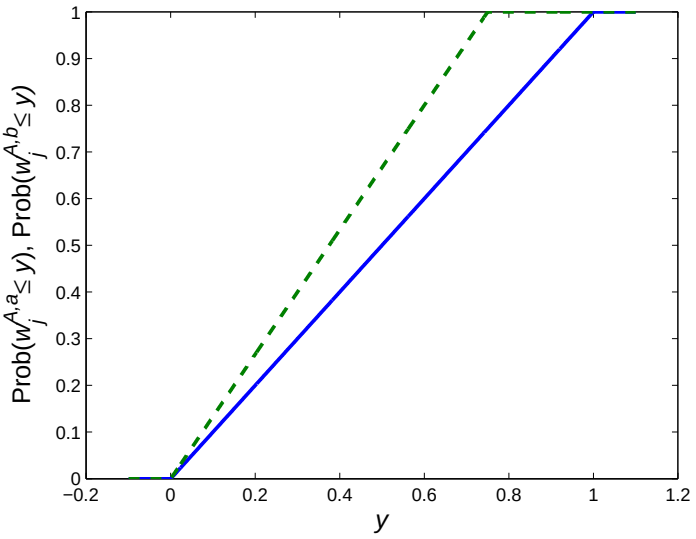

Figure A.5: The cumulative distribution functions of $w_j^{A,a}$ (solid line) and $w_j^{A,b}$ (dashed line) for $x = 0.75$.

can define the tail distribution $H_\varepsilon(y) := \mathbb{P}(w_j^{A,a} - w_j^{A,b} \geq y)$ by

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

By symmetry, the probability $\mathbb{P}(w_j^{B,b} - w_j^{B,a} \geq y)$ that the difference in values for the B-player is at least y is also given by $H_\varepsilon(y)$. So, we can identify $w_j^{A,a} - w_j^{A,b}$ and $w_j^{B,b} - w_j^{B,a}$ with the same random variable, denoted Δ_j, with tail distribution $H_\varepsilon(\cdot)$; see Figure A.6.

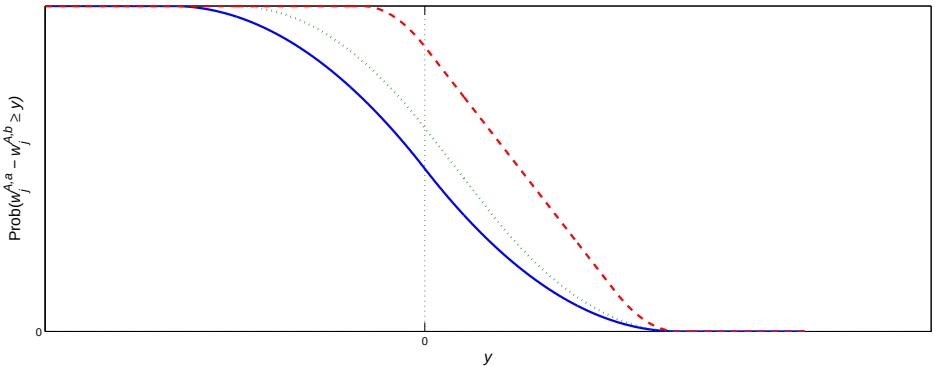

Figure A.6: The probability that $w_j^{A,a} - w_j^{A,b}$ is at least y, as a function of y, for $\varepsilon = 0$ (solid line); $\varepsilon = 0.125$ (dotted line); and $\varepsilon = 0.375$ (dashed line).

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.

Appendix A.2. Project values

We calculate the total value that players derive from their projects. Let $\widetilde{\Pi}(p)$ be the total project value for players that are assigned to project a when a proportion p of players with the strongest intrinsic preference for the group-preferred project are assigned to that project. That is, $\widetilde{\Pi}(p)$ is the sum (i.e., integral) of the values $w_j^{A,a}$ of the players j in group A that belong to the proportion p of 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 proportion $1-p$ of B-players with the strongest intrinsic preference for project a. As groups are symmetric, $\widetilde{\Pi}(p)$ is also equal to the total value derived from project b. So, $\Pi(h) = 2\widetilde{\Pi}(p)$ when the level of homophily is $h = p - \frac{1}{2}$. The next result characterizes the payoff $\widetilde{\Pi}(p)$ derived from a project.

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

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

Proof. To calculate total project value $\widetilde{\Pi}(p)$, fix a group, say A. Recall that in any social optimum, the proportion 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 cutoff y are assigned to project a, and the other A-players are assigned to project b. The proportion 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 defined in Appendix A.1. Since this tail distribution has different regimes, depending on y, we need to consider different cases. Rather than considering different ranges for the cutoff y, we will work with different ranges for $p = H_\varepsilon(y)$ as this turns out to be more convenient.

Case 1: $p \in [\frac{1}{2}, \frac{1}{2} + \varepsilon)$.. First suppose that the proportion p of players assigned to the 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.1) the equation reduces to $p = 1 - \frac{x}{2} - y$, or, equivalently,

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

For a given $y = y(p)$, if every A-player 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

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

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

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

The second term is for A-players for whom $w_j^{A,a}$ is so small (relative to the cutoff 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-players for whom $w_j^{A,a} - y \geq 0$. Working out the integrals and summing the terms gives the expression for $\widetilde{\Pi}(p)$ in the lemma for $p \in [\frac{1}{2}, \frac{1}{2} + \varepsilon)$.

Case 2: $p \in [\frac{1}{2}+\varepsilon, 1]$. . Next suppose $p \in [\frac{1}{2}+\varepsilon, 1]$. Again, fix a group, say A, and note that the A-players for whom $w_j^{A,a} - w_j^{A,b}$ exceeds a cutoff $z = z(p)$ 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

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

It will be convenient to work with a nonnegative cutoff, so define $y := -z \geq 0$. Then, rewriting gives[20]

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

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

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

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

Appendix B. Equilibrium analysis

We compare the outcomes predicted by the introspective process to the standard equilibrium prediction. 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 (i.e., impulse), 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 [cf. 38]. The game has more correlated 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 (e.g., choose s^2 if and only the signal prescribes s^1). Given this, there is a plethora of equilibria for the extended game.

We thus restrict attention to correlated equilibria in anonymous strategies. That is, 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. Furthermore, in the coordination stage, we restrict attention to correlated 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^m having received signal m is a best response if and only if

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

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

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

So, it remains to consider the matching stage. Suppose that $m^{A,a}$ and $m^{B,b}$ are the proportions of A-players and B-players who 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

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

or, equivalently,

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

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

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

In equilibrium, we must have that

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

Because the random variables $w_j^{A,a} - w_j^{A,b}$ and $w_j^{B,b} - w_j^{B,a}$ have the same distribution (cf. Appendix A.1), 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.1), the equilibrium condition reduces to

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

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

It is easy to see that the introspective equilibrium characterized in Proposition 3.3 is an equilibrium. However, the game has more equilibria. The point $p = 0$ is a fixed point of (B.1) if and only if $\beta \geq 1$. In an equilibrium with $p = 0$, all A-players adopt project b, even if they have a strong intrinsic preference for project a, and analogously for B-players. In this case, the incentives for interacting with the own group, measured by β, are so large that they dominate any intrinsic preference.

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

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

Importantly, the multiplicity of equilibria in the standard setting makes it difficult to derive unambiguous comparative statics. This is because as parameters are adjusted, the set of equilibria changes. Consider, for example, the effect of increasing the within-group similarity. As any introspective equilibrium is a correlated equilibrium, there is a correlated equilibrium where greater within-group similarity leads to more homophily (Corollary 3.4). But, varying the within-group similarity also changes the set of correlated equilibria. It is not hard to construct examples where increasing the within-group similarity gives rise to new (anonymous) correlated equilibria with lower levels of homophily.

Appendix C. Signaling and markers

Thus far, we have assumed that players sort by choosing projects. An alternative way to seek out similar others is by signaling. Here, we assume that players can use markers, that is, observable attributes such as tattoos, to signal which group they belong to.

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 that 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.[21] Each player is matched exactly once.[22] Players' decision to propose or to accept a proposal may depend on marker choice, but does not depend on players' identities or group membership, which is unobservable. If player j proposed to player j', and j' accepted j's proposal, then they play the coordination game in Section 2; if j's proposal was rejected by j', both get a payoff of zero.

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

We can prove an analogue of Proposition 3.3 for this setting:

Proposition Appendix C.1. [Equilibrium Characterization Marker Choice] *There is a unique introspective equilibrium of the extended game. In the unique equilibrium, players follow their impulse in the*

[21] 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.

[22] Such a matching is particularly straightforward to construct when there are finitely many players. Otherwise, we can use the matching process of Alós-Ferrer [4]. 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.

coordination game, and players' marker choices give rise to complete segregation ($h = \frac{1}{2}$) if and only if

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

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

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

In all cases, the fraction of players choosing the group-preferred marker exceeds the initial level (i.e., $h > h^0$).

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

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

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

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

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

The result now follows from the definition of the tail distribution $H_\varepsilon(\cdot)$ (Appendix A.1). □

This result shows that equilibrium takes a similar form as when players can sort by choosing projects. The equilibrium level of homophily is always higher than the level of homophily based on preferences over markers. The comparative statics are the same as before. For example, if the marginal benefit of interacting with the own group is sufficiently high, then there is full segregation.

So, even if players cannot influence the probability of meeting similar others by locating in a particular neighborhood or joining a club, they can nevertheless associate primarily with members of their own group if they can signal which group they belong to. This helps explain why groups are often marked by seemingly arbitrary traits such as tattoos or distinctive types of dress. Unlike in classic models of costly signaling, adopting a certain marker is *not* inherently more costly for one group than for another. Instead, the difference in signaling value of the markers across groups is endogenous in our model.

Appendix D. Proofs

Appendix D.1. Proof of Proposition 2.1

At level 0, all players follow their impulse. At level 1, a player with an impulse to choose action $s = s^1, s^2$ assigns probability

$$\hat{p}\cdot Q_{in} + (1-\hat{p})\cdot Q_{out}$$

to the other player having the same impulse (where $\hat{p}$ is the probability that a player is matched with a member of his own group). Since $\hat{p} > 0$ and $Q_{in} > \frac{1}{2}, Q_{out} \geq \frac{1}{2}$, this probability is strictly greater than $\frac{1}{2}$. It follows that the player's expected payoff at level 1 of following his impulse is strictly greater than $\frac{1}{2} \cdot v$; and the expected payoff of choosing the other action is strictly less than $\frac{1}{2} \cdot v$. Hence, at level 1, all players follow their impulse. A simple inductive argument then shows that at each level k, all players follow their impulse. □

Remark Appendix D.1. One might be concerned about the assumption in Section 2 that players have a positive probability to interact with members of their own group (i.e., $\hat{p} > 0$), given that $\hat{p}$ is an equilibrium outcome of the project choice game in Section 3. However, this point is not critical: by definition, if $\hat{p} = 0$, then the set of players that have zero probability of interacting their own group has measure 0. So, all our results go through unchanged, except that, in the knife-edge case where impulses are completely uninformative about the impulses of members of the other group (i.e., $Q_{out} = \frac{1}{2}$), the equilibrium in Proposition 2.1 is unique for "almost all" players (i.e., a set of players with measure 1) instead of for all players. ◁

Appendix D.2. Proof of Lemma 3.2

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 the notation $x := 1 - 2\varepsilon$ from Appendix A.1.

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

At higher levels, players choose projects based on their intrinsic values for the project as well as the coordination payoff they expect to receive at each project. Suppose that a share p_{k-1}^a of players with project a belong to group A, and likewise for project b and group B. Then, the probability that an A-player with project a is matched with a player of the own group is p_{k-1}^a, and the probability that a B-player with project a is matched with a player of the own group is $1 - p_{k-1}^a$. 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

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

This inequality can be rewritten as

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

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

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

where we have used the expression for the tail distribution $H_\varepsilon(y)$ from Appendix A.1. The same law of motion holds, of course, if a is replaced with b and A is replaced with B.

Fix a project π. We claim that for all k, $p_k^\pi \geq p_{k-1}^\pi$ and that the sequence $\{p_k^\pi\}_k$ is bounded. Then, by the monotone sequence convergence theorem, the limit p^π exists. To show this, fix $k > 0$. By the argument above,

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

where we have used the expression for the tail distribution $H_\varepsilon(y)$ from Appendix A.1. If $(2p_{k-1}^\pi - 1) \cdot \beta \geq 1 - 2\varepsilon$, the result is immediate, so suppose that $(2p_{k-1}^\pi - 1) \cdot \beta < 1 - 2\varepsilon$.

Define[23]

$$f(p) := 1 - \tfrac{1}{2(1-2\varepsilon)} \cdot (1 - 2\varepsilon - (2p - 1) \cdot \beta)^2,$$

and note that $f(\cdot)$ is (strictly) increasing in p. It is easy to check that $f(p_0^\pi) \geq p_0^\pi$ and that, for p such that $(2p^\pi - 1) \cdot \beta < 1 - 2\varepsilon$, $f(p) \leq 1$. Hence, for all $k > 0$, $f(p_{k-1}^\pi) \geq p_{k-1}^\pi$ and the limit p^π exists. Clearly, this argument does not depend on π, so we have $p^a = p^b$. □

Appendix D.3. Proof of Proposition 3.3

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 3.2, 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.1,

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

Equation (D.2) has two roots,

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

and

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

where we have used the notation $x := 1 - 2\varepsilon$. 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 $\beta > 0$.□

[23] We thank Pooya Ravari for this proof.

Appendix D.4. Proof of Corollary 3.4

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

Appendix D.5. Proof of Lemma 4.1

Let $p \in [\frac{1}{2}, 1]$ be the proportion 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

$$v \cdot [p \cdot Q_{in} + (1-p) \cdot Q_{out}];$$

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

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

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

$$v \cdot [(1-p) \cdot Q_{in} + p \cdot Q_{out}];$$

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

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

Adding all terms together gives the total coordination payoff for project a. A similar calculation, of course, applies to project b. □

Appendix D.6. Proof of Proposition 4.2

Maximizing social welfare is equivalent to maximizing social welfare per project. Per-project welfare is given by

$$\widetilde{W}(p) = v \cdot [Q_{in} \cdot \left(p^2 + (1-p)^2\right) + 2 \cdot p \cdot (1-p) \cdot \tfrac{1}{2}] + \widetilde{\Pi}(p),$$

where the first term are the per-project coordination payoffs (Lemma 4.1), and the second term is the value derived from a project (Lemma Appendix A.1). While the marginal benefit β of interacting with the own group is positive in the benchmark model, it may be negative if there are skill complementarities across groups (Section 4.2). We thus characterize the socially optimal level of homophily for arbitrary β. We prove the result by considering the first-order and second-order conditions. As in the proof of Lemma Appendix A.1, 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

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

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

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

and

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

It is straightforward to verify that $r_2 \leq \frac{1}{2}$ 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:

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

and

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

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 $h^* = r_1' - \frac{1}{2} > \varepsilon$. □

Appendix D.7. Proof of Proposition 4.4

We first characterize the equilibrium level of homophily when there are skill complementarities across groups. We then characterize the socially optimal level of homophily.

As before, let p_k^a be the probability that A-players at project a are matched with their own group at level k, and let p_k^b be the probability that B-players at project b are matched with their own group at level k. The limit is well-defined:[24]

Lemma Appendix D.2. *Suppose skill complementarities are sufficiently strong (i.e., $\beta < 0$) and that $\beta > -\frac{1}{2}$. The sequence $p_0^\pi, p_1^\pi, \ldots$ has a unique limit p^π for each project $\pi = a, b$. Moreover, the limits for the two projects coincide: $p^a = p^b$.*

Proof. Suppose $\beta \in (-\frac{1}{2}, 0)$. By an argument similar to the one in the proof of Lemma 3.2, it follows that the sequence $\{p_k^\pi\}_k$ is weakly decreasing and bounded for every project π. Moreover, $p_k^\pi < \frac{1}{2} + \varepsilon$ for all k. Again, by the monotone convergence theorem, the sequences $\{p_k^a\}_k$ and $\{p_k^b\}_k$ converge to a common limit p. □

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

Proposition Appendix D.3. [Skill Complementarities: Equilibrium] *Suppose $\beta \in (-\frac{1}{2}, 0]$. There is a unique introspective equilibrium. In the coordination game, players follow their impulse, and the equilibrium fraction of players choosing the group-preferred project is strictly below the initial level (i.e., $h < \varepsilon$), and is given by*

$$h = \frac{\varepsilon}{1-2\beta} > 0.$$

Proof. By Lemma Appendix D.2, $p_k^\pi \leq p_{k-1}^\pi$ for all k and $\pi = a, b$. By the monotone sequence convergence theorem, $p = \inf_k p_k^\pi$ (for arbitrary π). As before, we can find p by solving the fixed-point equation

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

Writing $y := -(2p-1)\cdot\beta$, we now need to consider two regimes: $y \in (0, \varepsilon)$ and $y \in [2\varepsilon, 1]$ (cf. Appendix A.1). In the latter 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

[24] It can be checked that if $\beta < -\frac{1}{2}$, then the sequence $p_0^\pi, p_1^\pi, \ldots$ does not settle down.

that lie outside the domain $(0, 2\varepsilon)$. So consider the former regime, where $H_\varepsilon(y) = 1 - \frac{1}{2}x - y$. The fixed-point equation has a unique solution y^*, with corresponding limiting probability

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

It can be checked that p is increasing in β, and lies in $(\frac{1}{2}, \frac{1}{2} + \varepsilon)$ for $\beta < 0$. □

So, skill complementarities across groups can reduce the equilibrium level of homophily. Our model shows how economic and cultural factors interact: if groups are similar in the sense that the within-group similarity index Q_{in} is close to the cross-group similarity index Q_{out}, then complementarities of skills become more important in shaping interactions. The following claim, which is proven as part of Proposition 4.2, characterizes the socially optimal level of homophily in the presence of skill complementarities.

Claim Appendix D.4. [Skill Complementarities: Social Optimum] Suppose $\beta < 0$. Full segregation is never optimal. The socially optimal level of homophily is given by:

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

The fraction of players choosing the group-preferred project in the social optimum is below the initial level (i.e., $h^* < h^0$). ◁

The proof is part of the proof of Proposition 4.2. Comparing this claim with Proposition Appendix D.3, we see that if skill complementarities are sufficiently strong, there can be too much homophily in equilibrium. □

References

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

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

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

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

[5] Alsan, M., O. Garrick, and G. C. Graziani (2018). Does diversity matter for health? Experimental evidence from Oakland. Working paper, NBER.

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

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

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

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

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

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

[12] Bauer-Wolf, J. (2018). Random roommates only. *Inside Higher Education*.

[13] Berry, J. W., Y. H. Poortinga, M. H. Segall, and P. R. Dasen (Eds.) (2002). *Cross-cultural psychology: Research and applications*. Cambridge University Press.

[14] 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.

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

[16] Buono, A. F. and J. L. Bowditch (2003). *The Human Side of Mergers and Acquisitions: Managing Collisions Between People, Cultures, and Organizations*. Beard Books.

[17] Butow, P., M. Bell, D. Goldstein, M. Sze, L. Aldridge, S. Abdo, M. Mikhail, S. Dong, R. Iedema, R. Ashgari, R. Hui, and M. Eisenbruch (2011). Grappling with cultural differences: Communication between oncologists and immigrant cancer patients with and without interpreters. *Patient Education and Counseling 84*, 398–405.

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

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

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

[21] 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.

[22] 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.

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

[24] Currarini, S. and F. Mengel (2016). Identity, homophily, and in-group bias. *European Economic Review 90*, 40–55.

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

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

[27] Fang, H. and G. C. Loury (2005). "Dysfunctional identities" can be rational. *American Economic Review 95*(2), 104–111.

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

[29] 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.

[30] Hall, J. A., J. T. Irish, D. L. Roter, C. M. Ehrlich, and L. H. Miller (1994). Gender in medical encounters: An analysis of physician and patient communication in a primary care setting. *Health Psychology 13*(5), 384–392.

[31] Heinke, M. and W. R. L. Louis (2009). Cultural background and individualistic-collectivistic values in relation to similarity, perspective taking, and empathy. *Journal of Applied Social Psychology 39*, 2570–2590.

[32] Hong, L. and S. E. Page (2001). Problem solving by heterogeneous agents. *Journal of Economic Theory 97*, 123–163.

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

[34] Huston, T. and G. Levinger (1978). Interpersonal attraction and relationships. *Annual Review of Psychology 29*, 115–156.

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

[36] Jackson, M. O., B. W. Rogers, and Y. Zenou (2017). The economic consequences of social-network structure. *Journal of Economic Literature 55*, 49–95.

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

[38] Kets, W. and A. Sandroni (2015). A theory of strategic uncertainty and cultural diversity. Working paper.

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

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

[41] Kreps, D. M. (1990b). *Game Theory and Economic Modelling*. Clarendon Lectures in Economics. Oxford University Press.

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

[43] Larsson, R. and M. Lubatkin (2001). Achieving acculturation in mergers and acquisitions: An international case survey. *Human Relations 54*, 1573–1607.

[44] Lau, D. C. and J. K. Murnighan (1998). Demographic diversity and faultlines: The compositional dynamics of organizational groups. *Academy of Management Review 23*, 325–340.

[45] Lazear, E. P. (1999). Globalisation and the market for team-mates. *Economic Journal 109*, C15–C40.

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

[47] McFarland, D. A., J. Moody, D. Diehl, J. A. Smith, and R. J. Thomas (2014). Network ecology and adolescent social structure. *American Sociological Review 79*, 1088–1121.

[48] 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.

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

[50] Morris, S. and H. S. Shin (2002). Social value of public information. *American Economic Review 92*, 1521–1534.

[51] Nelson, D. W. and R. Baumgarte (2004). Cross-cultural misunderstandings reduce empathic responding. *Journal of Applied Social Psychology 34*, 391–401.

[52] 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.

[53] 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.

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

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

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

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

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

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

[60] Williams, H. M., S. K. Parker, and N. Turner (2007). Perceived dissimilarity and perspective taking within work teams. *Group and Organizational Management 32*, 569–597.

[61] Young, H. (1998). *Individual Strategy and Social Structure: An Evolutionary Theory of Institutions*. Princeton University Press.