Community Leaders and the Preservation of Cultural Traits: Online Appendix

Anja Prummer Jan-Peter Siedlarek

November 28, 2016

1 Speed of Convergence

The speed of convergence in case no leader is present is determined by ability and the network structure. This is highlighted in the following proposition.

Proposition 12. *Consider a symmetric setting with k group members that assign weight $\gamma \in \left[\frac{1}{k}, 1\right]$ to their own identity and equal weight $(1-\gamma)/(k-1)$ to each of the $k-1$ other members.*

1. *If all group members have the same initial integration level, then group members of high ability integrate faster than those of lower ability, in the sense that at any point of time their identity is lower than that of group members of lower ability.*
2. *If group members have the same ability, then group members with lower initial identity integrate faster than those with higher initial identity, in the sense that at any point of time their identity is lower than that of group members with higher initial identity.*

Proof of Proposition 12 – Speed of Convergence

We first prove the following result concerning the ranking of identities in subsequent time periods. Proposition 12 then follows immediately.

Lemma. *Consider a symmetric setting with k group members that all assign weight $\gamma \in [\frac{1}{k}, 1]$ to their own identity and equal weight $(1-\gamma)/(k-1)$ to each of the $k-1$ other members.*

Consider any pair of group members i, j such that $\alpha_i \geq \alpha_j$. If $p_i^t \leq p_j^t$ this implies $p_i^{t+1} \leq p_j^{t+1}$. If the inequality on α or on p^t or on both is strict, then $p_i^{t+1} < p_j^{t+1}$.

We show that the condition $\gamma \in [\frac{1}{k}, 1]$ is sufficient for the result. Take group members i, j such that $\alpha_i \geq \alpha_j$ and let $p_i^t \leq p_j^t$. These assumptions imply $\hat{g}(p_i^t; \alpha_i) \geq \hat{g}(p_j^t; \alpha_j)$. Now, the difference between next period identities $p_i^{t+1} - p_j^{t+1}$ can be simplified to

$$\begin{aligned}
& p_i^{t+1} - p_j^{t+1} \\
&= \left[1 - \hat{g}(p_i^t; \alpha_i)\right] \left\{ \gamma p_i^t + \frac{1-\gamma}{k-1} \sum_{l \neq i} p_l^t \right\} - \left[1 - \hat{g}(p_j^t; \alpha_j)\right] \left\{ \gamma p_j^t + \frac{1-\gamma}{k-1} \sum_{l \neq j} p_l^t \right\} \\
& \left[\hat{g}(p_j^t; \alpha_j) - \hat{g}(p_i^t; \alpha_i)\right] \left\{ \frac{1-\gamma}{k-1} \sum_{l \neq i,j} p_l^t \right\} \\
& \quad + \left\{ \left[1 - \hat{g}(p_i^t; \alpha_i)\right] \left[\gamma p_i^t + \frac{1-\gamma}{k-1} p_j^t \right] - \left[1 - \hat{g}(p_j^t; \alpha_j)\right] \left[\gamma p_j^t + \frac{1-\gamma}{k-1} p_i^t \right] \right\}
\end{aligned}$$

The first part of this sum is non-positive given $\hat{g}(p_j^t; \alpha_j) - \hat{g}(p_i^t; \alpha_i) \leq 0$. Furthermore, the second part consists of the difference between two weighted averages of p_i^t and p_j^t. If $\gamma \geq 1/k$, then $\gamma \geq \frac{1-\gamma}{k-1}$ so that the first of these averages is weakly smaller than the second. In addition $\hat{g}(p_j^t; \alpha_j) - \hat{g}(p_i^t; \alpha_i) \leq 0$ implies $\left[1 - \hat{g}(p_i^t; \alpha_i)\right] \leq \left[1 - \hat{g}(p_j^t; \alpha_j)\right]$ which ensures that also the second part of the sum is negative. The sum of two non-positive numbers is non-positive and thus overall $p_i^{t+1} - p_j^{t+1} \leq 0$ as required. The result for strict inequality follows from recognizing that $\hat{g}(p_j^t; \alpha_j) - \hat{g}(p_j^t; \alpha_j) < 0$ if either $\alpha_i > \alpha_j$ or $p_i^t < p_j^t$ or both. In that case, both parts of the sum are strictly negative and thus $p_i^{t+1} - p_j^{t+1} < 0$.

Further, the differences in identity are also maintained for periods more than one step apart and we then arrive at the result stated in Proposition 12.

2 Network Structure and Ability

The steady state identity vector for group members is given by

$$\overline{p} = \left\{ \mathbf{I} - (\mathbf{I} - \mathbf{\Lambda}) \left[\mathbf{I} - \hat{\mathbf{G}}(\overline{p}) \right] \mathbf{\Gamma} \right\}^{-1} \; [\mathbf{I} - \hat{\mathbf{G}}(\overline{p})] \mathbf{\Lambda} \mathbf{1} p_L.$$

Note first that in the second half of the expression the leader's identity p_L is multiplied by $[\mathbf{I} - \hat{\mathbf{G}}(\overline{p})]\mathbf{\Lambda}$. Thus, for each group member i, this term depends on the influence the leader has on that member, λ_i, as well as the weight the group member assigns to the community as a whole relative to the host society, $[1 - \hat{g}(\overline{p}_i, \alpha_i)]$. Crucial for this weight is the productivity α_i, which leads higher ability group members to assign a lower weight to the leader. As this part depends only on a group member's individual characteristics and is independent of other group members, we define this half of the expression as the $n \times 1$ vector of "idiosyncratic" identities. The network structure matters in the first part of the previous equation, though. Rewriting yields

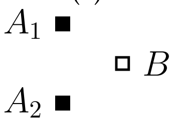
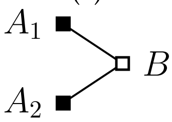
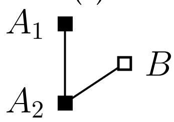
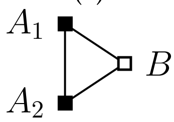
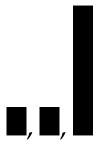
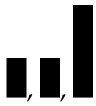
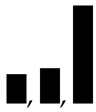
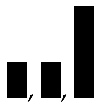
$$\left\{\mathbf{I} - (\mathbf{I} - \mathbf{\Lambda}) \left[\mathbf{I} - \hat{\mathbf{G}}(\overline{p})\right] \mathbf{\Gamma}\right\}^{-1} = \sum_{k=0}^{\infty} \left\{(\mathbf{I} - \mathbf{\Lambda}) \left[\mathbf{I} - \hat{\mathbf{G}}(\overline{p})\right]\right\}^{k} \mathbf{\Gamma}^{k} \tag{1}$$

Each element of the sum consists of two components. First, $\mathbf{\Gamma}^k$ reflects the influence weights among the group after k iterations, that is, element i, j of $\mathbf{\Gamma}$ to the k-th power gives the relative influence that the period t identity of group member j has on the period $t + k$ identity of group member i. For $k > 1$ this matrix also captures indirect influences and allows group members to influence each other that are not directly connected in the influence network. For example, with $k = 2$, element i, j of $\mathbf{\Gamma}^k$ would be positive if j is at most 2 steps away from i in the influence network. The second component $(\mathbf{I} - \mathbf{\Lambda}) \left[\mathbf{I} - \hat{\mathbf{G}}(\overline{p})\right]$ is a diagonal matrix with elements on the diagonal $(1 - \lambda_i)\hat{g}(\overline{p}_i; \alpha_i) < 1$ that act as individual specific discount factors: the higher k, the smaller the factor that is multiplied with the term $\mathbf{\Gamma}^k$.

For each group member the expression above thus adds up the direct and indirect influences from across the group member network and weighs them in a way such that shorter network connections receive higher weight than longer ones. The infinite sum presents the relative influence weights in the limit when all possible influence paths through the group member network are accounted for. Overall, equation (1) can thus be seen as a matrix of influence weights that captures the long-run pattern of influence, once all direct and indirect influences amongst the group members have played out.

This matrix of long-run weights is then applied to the vector of idiosyncratic identity terms $[\mathbf{I} - \hat{\mathbf{G}}(\overline{p})]\mathbf{\Lambda}$ from above, that captures the identity level of each group member without network influence. The result is a vector of steady state identities that each present a weighted average of the idiosyncratic identities, with the weights given by the infinite sum expression in equation (1).

To illustrate the relationship expressed in Proposition 4 between characteristics of the group network and the resulting long run identities consider the following simple example involving three group members. Two group members, A_1 and A_2, have a high productivity α_A, the third group member B a lower productivity α_B, with $\alpha_A > \alpha_B$. We consider different network structures among the three. In each case, we focus on symmetric distributions of influence weights such that if one group member is connected to exactly n other group members, each neighbor and the group member himself receive weight $1/(n + 1)$ in the updating process. Table 1 shows the results.

	(1)	(2)	(3)	(4)
				
Degree (A_1, A_2, B)	$0, 0, 0$	$1, 1, 2$	$1, 2, 1$	$2, 2, 2$
Clustering Coefficient	0	0	0	1
Density	0	$2/3$	$2/3$	1
Long-Run Identities (A_1, A_2, B)	 (0.20, 0.20, 0.92)	 (0.28, 0.28, 0.66)	 (0.21, 0.25, 0.70)	 (0.25, 0.25, 0.65)

Note: *The example uses* $\hat{g}(p; \alpha) = \alpha e^{-0.9p}$. *With* $\alpha \leq 1$ *this function ensures* $\hat{g}(p; \alpha) \in (0, 1)$ *and in addition satisfies Assumption 1. Other parameters are as follows:* $p_L = 1$, *initial* $p_i^0 = 1 \ \forall \ i$, $\alpha_A = 0.8$ *and* $\alpha_B = 0.1$.

Table 1: Long-Run Identities as a Function of Network Structure

(1) Isolated Group Members In the isolated case, long-run identities for the productive group members A_1 and A_2 converge to a common level that is significantly lower than that of group member B.

(2) / (3) Open Triangle / Line In the open triangle, we distinguish two cases. In the first case in column (2), B is placed at the center, connecting to both A_1 and A_2, while in the second case in column (3), one of the high productivity group members is located at the center. We see that in both cases, the existence of connections between the group members attenuates the differences between the high and low productivity group members relative to the isolated structure. In the case with B at the center, A_1 and A_2 show the same long run identity and B's identity is lower than in the second case with B at an end node. In that second case, A_1 and A_2 will have different steady state identities, with the group member that is directly connected to B converging to an identity above that of the other group member.

(4) Closed Triangle / Circle In the case of the closed triangle, A_1 and A_2 once more display the same long run identity reflecting their symmetric positions in the group. Note that the extra connections attenuate the differences between high and low α group members. These are smaller than in both the isolated and the open triangle setting.

The examples illustrate a number of features of the long run identities as characterized in Proposition 4. First, individual attributes such as the ability parameter α_i and the individual specific leader influence λ_i are important drivers of differences in identity between group members. Second, connections in the influence network tend to move the long-run identities of the connected group members closer together.

In our example, the high productivity types A have a low identity when isolated but their identities become more pronounced when we add connections to the low productivity type, and vice versa for low productivity type B. In general, additional connections help attenuate the difference between the most extreme identities. A similar pattern emerges for clustering in groups. Comparing columns (2) and (4) in Table 1, we find that the difference between types A and B is smaller in the highly clustered closed triangle.[1]

In light of these results, we can also discuss the impact that removing individual group members has on the remainder of the group. As before, the impact of this intervention will depend on both idiosyncratic factors (such as ability and susceptibility to leader influence) and the extent of connectivity. First, a group member whose identity is more different from that of the rest of the group will have a larger effect than one who is more similar to others, holding all else equal. Second, if two group members have similar characteristics but vary in connectivity, then similar to the "key player" analysis of Ballester et al. (2006) the influence matrix also matters and the removal of a more connected connected player will have a greater impact than that of a less connected one.

Generally, the long run outcome of an individual group member depends on the interplay of both idiosyncratic factors and network structure. Holding social influence constant as in our example, higher ability types will integrate more and adopt a lower long-run identity than lower ability types. However, if the influence networks vary significantly between types, then it is possible for a higher ability group member to integrate less and have a higher identity in the long run than a lower ability group member.

3 Cutting Ties

We have assumed throughout that group members' connections to their own community are weakened as they become more integrated in the host society. However, in our baseline setting group members never fully leave their group and it seems plausible that agents can also cut their connection to the community once their identity, that is their identification with their own group, is sufficiently low. We consider this scenario, in which agents severe

[1]Note however that from this simple example we cannot disentangle the effect of clustering from the effect of the additional connection between A_1 and A_2.

ties to their community if they are sufficiently integrated, in what follows.

Note first that in the setting with two groups, there are no incentives for the group leader to unilaterally exclude one group from the community.

Proposition 13. *In the two-type setting with $p^{\max}$ sufficiently high, the leader never excludes one group from the community.*

The proof of this result can be found at the end of this section.

However, group members might still leave according to their own wishes. Suppose there exists a boundary identity, $\underline{p}$, such that all group members with an identity lower than $\underline{p}$ sever ties to the group and only members with a higher identity remain in the community. How does the leader respond to the possibility of group members cutting ties?

The problem is trivial if there is no group member whose identity lies below the bound at the optimum: the leader is then effectively unconstrained and simply behaves as if the bound were not present. However, if a group member's identity at the unconstrained maximum lies below the bound, which induces this group member to leave the community, the leader has two possibilities to respond. First, he can accommodate the group members below the bound by increasing his own identity which in turn increases the identities of the group members until no group member has the incentive to leave the community. Alternatively, he allows for the group members to leave and focuses on the remaining group members, selecting a possibly different level of p_L that is optimal for the smaller group. In order to calculate which option is more beneficial, we need to compare the payoff from excluding a group to the payoff from adjusting the identity to retain the group.

To illustrate this comparison, we focus on the case of two types, A and B, with symmetric influence pattern where as before we assume that group members of type A have a higher ability than group members of type B, such that in steady state $\overline{p}_A < \underline{p} < \overline{p}_B$. Denote the share of group members of type A by κ. We index variables with a superscript B to indicate that only members of type B belong to the group, and by AB if both types remain in the group. Then, Π^B, that is, the payoff if only type B remains in the group, is given by

$$\Pi^B = (1-\kappa)\,\pi(\overline{p}_B^B, \alpha_B f(\overline{H}_B^B))$$

If both types remain in the group, then $\overline{p}_A^{AB} = \underline{p}$. We denote B's identity in this case by $\overline{p}_B^{AB}$.

Then, the payoff is given by

$$\Pi^{AB} = \kappa\pi(\underline{p}, \alpha_A f(\overline{H})) + (1-\kappa)\,\pi\left(\overline{p}_B^{AB}, \alpha_B f\left(\overline{H}_B^{AB}\right)\right)$$

It is straightforward to see that the smaller κ, the more likely it is that the leader does not increase his identity to retain members of type A. Depending on the exact specification it can either be optimal for the leader to retain A types or to let them go. We construct an example for each of these outcomes for a religious leader with the functional forms shown in Table 2.

$f(H)$	$g(H)$	$c(p)$
$kH - \frac{1}{6}H^3$	$\frac{H}{1+H}$	$\frac{1}{2}(1+p)$

Table 2: Functional Forms

Then, we can provide an example for which $\Pi^B < \Pi^{AB}$ and one for which the reverse holds. It is better to keep both group members in the community, $\Pi^B < \Pi^{AB}$ for the parameters specified in the first row of Table 3, whereas the second row shows the parameters under which it is optimal to omit type A. Based on these parameters we can show what the

	κ	β	λ	γ	α_A	α_B	k	$\underline{p}$
(1) $\Pi^B < \Pi^{AB}$	1/2	1/4	1/2	1/2	3/4	1/2	4	3/4
(2) $\Pi^B > \Pi^{AB}$	1/200	2/5	1/2	1/2	2	1/2	3	9/10

Table 3: Example

steady state identities, the identities of the leader and earnings of the group members are, respectively. This is summarized in Table 4, where again the first row gives the case when it is better to keep group member A within the community, whereas in the second row it is better if A has severed his ties.

	p_L^B	$\overline{p}_B^B$	$\alpha_B f(\overline{H}_B^B)$	Π^B	p_L^{AB}	$\overline{p}_B^{AB}$	$\alpha_A f(\overline{H}_A^{AB})$	$\alpha_B f(\overline{H}_B^{AB})$	Π^{AB}
(1)	3.93	.75	3.45	1.38	4.29	.82	5.46	3.42	3.52
(2)	3.70	.95	1.92	1.52	4.70	1.39	9.70	1.55	1.50

Table 4: Identities and Earnings

Note that in the first case where it is optimal to retain group A, the leader would prefer to set a lower identity than 3.93 if he only faces B types. However, he cannot do so as then the group members of type A would leave the group. What can be noted here is that whenever only type B's remain in the group, the leader sets a lower identity, which results in better integration and higher earnings of the group members. Therefore, A types leaving the group leads to better integration outcomes through two channels: the A types will integrate completely and additionally, the B types will integrate better.

It is natural to ask to what extent the effect of a policy measure in the case where ties vary only in their intensity carry over to the setting where ties can be cut. To see this more clearly, suppose B's productivity α_B decreases, from $1/2$ to $1/3$. We focus here on example (1) as such a decrease in α_B in example (2) induces the leader to choose maximal cultural distinction (independently of whether the boundaries are present or not). The identities and earnings if ties only change continuously are given in the first row of Table 5. The second row shows the corresponding value if α_B decreases from $1/2$ to $1/3$.

	p_L	$\overline{p}_A$	$\overline{p}_B$	$\alpha_A f(\overline{H}_A)$	$\alpha_B f(\overline{H}_B)$	Π
(1) with $\alpha_B = 1/2$	4	.70	.76	5.47	3.44	3.52
(1) with $\alpha_B = 1/3$	1.45	.24	.28	5.56	2.25	2.99

Table 5: Identities and Earnings without Cutting Ties

This shows that if α_B decreases, the leader lowers his identity significantly. Both types of identities fall and while the earnings of type A increase, those of type B decrease.

If ties to the community are cut for sufficiently low identity, then the payoffs for $\alpha_B = 1/3$ are as given in Table 6.

	p_L^B	$\overline{p}_B^B$	$\alpha_B f(\overline{H}_B^B)$	Π^B	p_L^{AB}	$\overline{p}_B^{AB}$	$\alpha_A f(\overline{H}_A^{AB})$	$\alpha_B f(\overline{H}_B^{AB})$	Π^{AB}
(1)	3.93	.75	1.96	.83	4.15	1.10	5.46	1.61	2.88

Table 6: Identities and Earnings: $\alpha_B = 1/3$

If only type B remains in the group, the leader has to set the same identity as before the change in α_B in order to satisfy the boundary condition. If both group members remain in the group, then the leader sets a lower identity similar to the case where ties cannot be cut. Given that group members B now have a higher identity for a given leader choice p_L, the leader can reduce the p_L he chooses in order to keep A within the group. Therefore, the policy implications in this case are exactly the same as if ties cannot be cut.

Generally, we consider the case in which all group members remain in the group to be the most interesting and realistic one. As can be seen from the example presented, the share of productive types needs to be very small in order for the productive group members to be omitted from the group.[2] If it is only beneficial to let group members leave, if they constitute a sufficiently small part, then policy makers are still faced with large groups of immigrants who will not integrate fully and essentially the problem remains the same as in the case where ties are not cut. In fact the problem is exacerbated, as group leaders will now choose a higher level of cultural distinction compared to the case where ties cannot be cut and the impact of the policies is more limited. Whereas the leader's identity fell dramatically when cutting links was not possible as a response to the decrease in α_B, the reduction in the leader's identity if ties can be eliminated is much more moderate.

Proof of Proposition 13: Leader Exclusion

Suppose there are two types of group members, A and B and that the share of type A members is given by κ. Consider without loss of generality the case where the leader excludes type A members and only group members of type B belong to the group. The payoff to the group leader is then given by

$$\Pi^B = (1-\kappa)\,\pi(\overline{p}_B^B, \alpha_B f(\overline{H}_B^B)),$$

where $\overline{p}_B^B$ is the type B steady state identity determined by the optimal leader identity in case the group only contains B types. We denote the corresponding leader identity by p_L^B. The steady state is therefore characterized by

$$\overline{p}_B^B = \left(1 - \hat{g}(\overline{p}_B^B; \alpha_B)\right)\left(\lambda_B p_L^B + (1-\lambda_B)\overline{p}_B^B\right)$$

Now consider the payoff if agents of type A are part of the community and the leader chooses his identity p_L^{AB} such that $\overline{p}_B^{AB} = \overline{p}_B^B$. With $p^{\max}$ sufficiently high, this is always possible by adjusting p_L^{AB} appropriately. Then the payoff is given by

$$\begin{aligned}
\Pi^{AB} &= \kappa\pi(\overline{p}_A^{AB}, \alpha_A f(\overline{H}_A^{AB})) + (1-\kappa)\,\pi(\overline{p}_B^{AB}, \alpha_B f(\overline{H}_B^{AB})) \\
&= \kappa\pi(\overline{p}_A^{AB}, \alpha_A f(\overline{H}_A^{AB})) + (1-\kappa)\,\pi(\overline{p}_B^{B}, \alpha_B f(\overline{H}_B^{B})) \\
&> (1-\kappa)\,\pi(\overline{p}_B^{B}, \alpha_B f(\overline{H}_B^{B}))
\end{aligned}$$

Therefore, the payoff when agents of type A belong to the group is higher than the payoff when they do not belong to the group. Thus, a leader does not have an incentive to omit

[2] This seems to be a general feature of the examples we considered.

one type from the group.

4 SKILL INVESTMENT WITH FORWARD-LOOKING AGENTS

We discuss optimal skill investment of group members with foresight in two scenarios based on different degrees of foresight: In Scenario 1, group members evaluate their skills investment on a forward-looking basis, considering in particular that skills may be durable. However they remain myopic about future identity changes. In Scenario 2, group members incorporate a prediction of their own future identity as well as the path of other group member identities in the skills investment decision.

Scenario 1: Group Members Ignore Identity Changes

Consider first a group member with starting identity p^0, ability α and discount factor ρ. The group member takes p^0 as given and fixed forever.

We denote the decay factor of skills by $\delta \in [0, 1)$ and generalize the cost function to $C(h, p)$ to bring out the impact of the linearity assumption made in the main part of the paper. The stock of skills H then acts as a state variable.

The problem of the group member is to maximize:

$$V(H^0) = \sum_{t=1}^{\infty} \rho^t \left[\alpha f(H^t) - C\left(H^t - \delta H^{t-1}, p^0\right)\right].$$

By the principle of optimality, we can rewrite the problem recursively using the Bellman equation

$$V(H) = \max_{H'} \left\{\alpha f(H') - C\left(H' - \delta H\right) + \rho V(H')\right\}.$$

As the period payoff function and the updating function are continuous and bounded we can analyze the problem using standard dynamic programming techniques that yield the Euler equation connecting current investment H, tomorrow's H' and for two periods ahead H'':

$$\alpha f'(H') - C'\left(H' - \delta H\right) = -\rho\delta C'\left(H'' - \delta H'\right) < 0.$$

The right hand side is negative as $C' > 0$. Thus, at the optimum a group member invests more today, anticipating that a higher H' tomorrow will yield additional benefits in the future by reducing the need to invest then. At the steady state $H = H' = H'' = \overline{H}$ and thus:

$$\alpha f'(\overline{H}) - C'\left((1-\delta)\overline{H}\right) = -\rho\delta C'\left((1-\delta)\overline{H}\right).$$

The term on the right is zero if $\rho = 0$ or $\delta = 0$. Thus, if skills decay fully and $\delta = 0$ then myopic outcome is equal to the forward looking one. However, if $\delta > 0$, then skills are durable and thus valuable in the future. As a consequence, a group member with $\rho > 0$ will invest more today than with $\rho = 0$. Steady state investment is higher under forward looking behavior than in the myopic case.

Scenario 2: Group Members Predict Identity Changes

In this section we analyze a variant of our model in which group members are forward-looking and anticipating both their future identity as well as the path of identities of other members of the group. We characterize a forward-looking equilibrium in which each group member solves a dynamic problem maximizing payoffs while taking as given the path of other group members.

Maximization Problem Denote by $\{H_i^t\}_{t=0}^{\infty}$ and $\{p_i^t\}_{t=0}^{\infty}$ the path of player i's skill level and identity, respectively, and similarly $\{p_{-i}^t\}_{t=0}^{\infty}$ for all other players. Group member i maximizes:

$$\max_{\{p_i^t\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \rho^t \left\{\alpha f(H_i^t) - \left(c_0 + c_1 p_i^t\right) H_i^t\right\}$$

subject to the updating equation:

$$p_i^{t+1} = \left[1 - g(H_i^t; \alpha_i)\right] \left\{\lambda p_L + (1-\lambda)\left[\gamma_{ii} p_i^t + \sum_{j \neq i} \gamma_{ij} p_j^t\right]\right\}$$

Player i takes the full path of identities for other players $\{p_{-i}^t\}_{t=0}^{\infty}$ as given and chooses $\{p_i^t\}_{t=0}^{\infty}$ accordingly. In equilibrium, all players will individually select a path that represents a best response to other group member's paths. This equilibrium reflects an *open loop* Nash equilibrium and can be interpreted as the Nash equilibrium of the game where players can commit to their future path at time zero.[3]

[3] For further discussion of open loop equilibria and related concepts see e.g. Section 13.4.1 in Fudenberg and

Effect of Additional Periods on Skills Investment To bring out the effect that extending the investment horizon has on group member investment decisions when they take account of changes in identity both of themselves and other group members, consider first a finite time problem with final period T. The myopic decision problem then corresponds to the final period in which group member i takes p_i^T and p_{-i}^T as given and maximizes:

$$\alpha_i f(H_i^T) - (c_0 + c_1 p_i^T) H_i^T$$

As shown in the main part of the paper, the solution is characterized by the first order condition

$$\alpha_i f'(H_i^T) - c_0 + c_1 p_i^T = 0$$

The optimal level of investment is independent of other group members' identities or investment levels.

Now, add one further time period by considering the problem from period $T-1$. Denote by $V_i^T(p_i^T)$ the maximum utility achieved by player i in period T. Note that in this notation $V^T(\cdot)$ explicitly keeps track of time as the problem is not stationary: the value of starting with identity level p_i in final period T is different from starting in period $T-1$. Furthermore the identities of other players are permitted to change and this means that until a stationary outcome in which identities of other players do not change is reached, we have to keep track of the time period and the corresponding identities for other players in our calculations.

Note further that the value function is decreasing in p_i^T: a lower identity reduces the marginal cost of skills in investment and thereby allows i to achieve a higher payoff from investment. The effect can be expressed using the envelope theorem:

$$\begin{aligned} V_i^T(p_i^T) &= \max_{H_i^T} \left\{ \alpha_i f(H_i^T) - (c_0 + c_1 p_i^T) H_i^T \right\} \\ (V_i^T)'(p_i^T) &= -\, c_1 H_i^T < 0 \end{aligned}$$

Consider now the problem from period $T-1$. The value function for the group member

Tirole (1991).

can then be written as:

$$V_i^{T-1}(p_i^{T-1}) = \max_{H_i^{T-1}} \left\{ \alpha_i f(H_i^{T-1}) - (c_0 + c_1 p_i^{T-1}) H_i^{T-1} + \rho V_i^T(p_i^T) \right\}$$

The first order condition for the choice of H_i^{T-1} now includes the effect it has on the starting identity for final period T via the transition equation.

$$\alpha_i f'(H_i^{T-1}) - (c_0 + c_1 p_i^{T-1}) H_i^{T-1} + \frac{\partial p_i^T}{\partial H_i^{T-1}} \rho (V_i^T)'(p_i^T) = 0 \tag{2}$$

where

$$\frac{\partial p_i^T}{\partial H_i^{T-1}} = -g'(H_i^{T-1}) \left[\lambda p_L + (1-\lambda) \sum_j \gamma_{ij} p_j^{T-1} \right] < 0$$

Compared to the myopic approach and the final period, this condition reflects an additional incentive to invest in H_i^{T-1} represented in the second term on the right hand side of equation (2): a higher H_i^{T-1} implies player i will enter the final period with a lower identity and this lower identity means a greater payoff in that final period. Thus, compared to the myopic approach where this effect is ignored, this leads to greater levels of skills and lower identity. The strength of the additional effect depends on the identities of all group members that player i is connected to. The higher the identities of those other group members, the more they push player i's identity up and therefore the greater the payoff from investing in H_i^{T-1} to shift influence away from the group and towards the host society.

Consider now the derivative of $V_i^{T-1}(p_i^{T-1})$. As before, by the envelope theorem we can focus on the direct effect of changes in p_i^{T-1} yielding the following expression:

$$(V_i^{T-1})'(p_i^{T-1}) = -\, c_1 H_i^{T-1} + \frac{\partial p_i^T}{\partial p_i^{T-1}} \rho (V_i^T)'(p_i^T) \tag{3}$$

where

$$\frac{\partial p_i^T}{\partial p_i^{T-1}} = \left[1 - g(H_i^{T-1}) \right] (1-\lambda) \gamma_{ii} > 0$$

The additional term captures the effect that a change in p^{T-1} has on next period identity through the identity updating equation.

Turning now to the problem as seen from period $T-2$, the first order condition takes account of this additional term as follows:

$$
\begin{aligned}
&\alpha_i f'(H_i^{T-2}) - (c_0 + c_1 p_i^{T-2})H_i^{T-2} + \frac{\partial p_i^{T-1}}{\partial H_i^{T-2}}\rho(V_i^{T-1})'(p_i^{T-1}) \\
=&\alpha_i f'(H_i^{T-2}) - (c_0 + c_1 p_i^{T-2})H_i^{T-2} + \frac{\partial p_i^{T-1}}{\partial H_i^{T-2}}\rho\left\{-c_1 H_i^{T-1} + \frac{\partial p_i^{T}}{\partial p_i^{T-1}}\rho(V_i^{T})'(p_i^{T})\right\} \\
=&0
\end{aligned}
$$

As both terms in curly braces are negative, there is thus an additional incentive for player i to choose a higher level of H_i^{T-2} relative to time period $T-1$: a higher skill level implies a lower next period identity that – in addition to reducing marginal cost of investment in period $T-1$ – will also lead to a lower identity in period T, adding further benefits. Note that the strength of this effect depends on the parameter γ_{ii}, i.e., the extent to which past identity influences next period identity for player i.

As we roll back the problem further in time, these benefits will accumulate further, so that in period 1 an increase in H_i^1 will reduce not only p_i^2 but also all future p_i^t. Note that overall the problem remains well defined due to concavity of $f(\cdot)$ and the linearity of marginal costs. There thus exists a unique solution of the sequence problem for any given path of other group members' identities over time.

Forward-Looking Skills Investment at the Steady State As noted above, for a given path of identities of other group members the optimization problem of group member i is not necessarily stationary across time. To be able to say more about impact of forward-looking behavior on long-run skills investment we consider a group which has reached a steady state.

In steady state $p_i^t = \overline{p_i}$, $H_i^t = \overline{H_i}$ as well as $V^t(p_i^t) = V(\overline{p_i})$ for all players and thus we have a stationary investment problem from the perspective of group member i. This allows dropping the superscript t from the value functions and the derivative of the value function in equation (3) can then be rearranged to yield

$$
\begin{aligned}
V'(\overline{p_i}) =& \frac{-c_1\overline{H_i}}{1-\rho(1-\lambda)\left[1-g(\overline{H_i})\right]\gamma_{ii}} \\
=& \left(-c_1\overline{H_i}\right)\sum_{t=0}^{\infty}\left[\rho(1-\lambda)\left[1-g(\overline{H_i})\right]\gamma_{ii}\right]^t
\end{aligned}
$$

The expression in terms of an infinite sum highlights the way in which an increase in identity p_i creates benefits in terms of marginal cost reductions proportional to $c_1 H_i$ both directly and indirectly via its effect on next period p_i and subsequent periods for the infinite future.

The first order condition for the optimal skills level in the steady state can be rearranged to

$$\begin{aligned}&\alpha_i f'(\overline{H_i}) - (c_0 + c_1\overline{p_i})\\&= -\rho\frac{c_1\overline{H_i}}{1-\rho(1-\lambda)\left[1-g(\overline{H_i})\right]\gamma_{ii}}g'(\overline{H_i})\left[\lambda p_L + (1-\lambda)\left(\gamma_{ii}\overline{p_i} + \sum_{j\neq i}\gamma_{ij}\overline{p_j}\right)\right]\end{aligned}$$

This condition shows that relative to the myopic optimization problem, the group members have an incentive to invest more in skills, resulting in lower steady state identities. This incentive derives from the negative relationship between skills and identity and the extent to which lower identity reduces the marginal costs of further investment, both directly, and indirectly through lowering future identities via the self-influence parameter γ_{ii}.

Implications for Convergence of Identities

The derivations in the previous sections bring out the effect of extending the planning horizon of group members in their decision how much to invest in skills. As we show, making players forward-looking tends to lead to greater investment levels than the myopic case, both in the setting with durable skills (Scenario 1) and in the setting with players predicting future paths for their own and other players' identities (Scenario 2). Through the function $g(\cdot)$ these higher investment levels will translate mechanically into an overall lower path of group member identities compared to the myopic case.

In terms of implications for the dynamics of our model, we can capture these differences by suitably adjusting the function $\hat{g}(p)$, which maps a group member's identity into the weight assigned to the host society and implicitly reflects the optimal choice of investment. With forward looking agents we would expect to see this function to increase for every p. However, as long as the resulting new function $\hat{g}(p)$ satisfies our Assumption 1, the convergence of the system overall is assured as before and Proposition 3 will go through. Furthermore, the qualitative properties of the long run outcome such as incomplete assimilation of group members remain unaffected.

5 Proofs

Proof of Proposition 13: Comparative Statics γ

Note that as $F_{AB} \equiv Y_A \frac{\partial \overline{p}_A}{\partial p_L} + Y_B \frac{\partial \overline{p}_B}{\partial p_L} = 0$, it follows that $Y_B = -Y_A \frac{\partial \overline{p}_A}{\partial p_L} \left(\frac{\partial \overline{p}_B}{\partial p_L}\right)^{-1}$. Based on this we can then turn to the two parts of the proof.

Part 1 We know that

$$\underbrace{\frac{\partial Y_A}{\partial \overline{p}_A}}_{\leq 0} \underbrace{\frac{\partial \overline{p}_A}{\partial \gamma}}_{<0} \underbrace{\frac{\partial \overline{p}_A}{\partial p_L}}_{>0} + \underbrace{\frac{\partial Y_B}{\partial \overline{p}_B}}_{\leq 0} \underbrace{\frac{\partial \overline{p}_B}{\partial \gamma}}_{>0} \underbrace{\frac{\partial \overline{p}_B}{\partial p_L}}_{>0}$$

The terms $\frac{\partial \overline{p}_i}{\partial p_L}$ and $\frac{\partial \overline{p}_i}{\partial \gamma}$ can be determined through the Implicit Function Theorem, where similar to before we define

$$|J| = \frac{\partial F_A}{\partial p_A}\frac{\partial F_B}{\partial p_B} - \frac{\partial F_A}{\partial p_B}\frac{\partial F_B}{\partial p_A}, \quad |J_i(p_L)| = \frac{\partial F_i}{\partial p_L}\frac{\partial F_{-i}}{\partial p_{-i}} - \frac{\partial F_i}{\partial p_{-i}}\frac{\partial F_{-i}}{\partial p_L}, \quad |J_i(\gamma)| = \frac{\partial F_i}{\partial \gamma}\frac{\partial F_{-i}}{\partial p_{-i}} - \frac{\partial F_i}{\partial p_{-i}}\frac{\partial F_{-i}}{\partial \gamma}$$

Then, $\frac{\partial \overline{p}_A}{\partial p_L} = -\frac{|J_A(p_L)|}{|J|}$, $\frac{\partial \overline{p}_B}{\partial p_L} = -\frac{|J_B(p_L)|}{|J|}$ and $\frac{\partial \overline{p}_A}{\partial \gamma} = -\frac{|J_A(\gamma)|}{|J|}$, $\frac{\partial \overline{p}_B}{\partial \gamma} = -\frac{|J_B(\gamma)|}{|J|}$.

We are interested in comparing the magnitudes of the different effects and in order to compare them we define

$$X_i \equiv 1 + \frac{\partial \hat{g}(\overline{p}_i;\alpha_i)}{\partial \overline{p}_i}\left(\lambda p_L + (1-\lambda)(\gamma \overline{p}_i + (1-\gamma)\overline{p}_j)\right) - (1-\lambda)\gamma\left(1-\hat{g}(\overline{p}_i;\alpha_i)\right) > 0$$

We first show that $\frac{\partial \overline{p}_A}{\partial p_L} < \frac{\partial \overline{p}_B}{\partial p_L}$ which is equivalent to showing that $-|J_A(p_L)| < -|J_B(p_L)|$.

$$-|J_A(p_L)| = X_B \lambda\left(1-\hat{g}(\overline{p}_A;\alpha_A)\right) + \lambda(1-\lambda)(1-\gamma)\left(1-\hat{g}(\overline{p}_A;\alpha_A)\right)\left(1-\hat{g}(\overline{p}_B;\alpha_B)\right) > 0$$
$$-|J_B(p_L)| = X_A \lambda\left(1-\hat{g}(\overline{p}_B;\alpha_B)\right) + \lambda(1-\lambda)(1-\gamma)\left(1-\hat{g}(\overline{p}_A;\alpha_A)\right)\left(1-\hat{g}(\overline{p}_B;\alpha_B)\right) > 0$$

Thus $-|J_A(p_L)| < -|J_B(p_L)|$ if

$$\begin{aligned} & X_B\left(1-\hat{g}(\overline{p}_A;\alpha_A)\right) < X_A\left(1-\hat{g}(\overline{p}_B;\alpha_B)\right) \\ \Leftrightarrow \quad & \left(1-\hat{g}(\overline{p}_B;\alpha_B)\right)\left(1 + \frac{\partial \hat{g}(\overline{p}_A;\alpha_A)}{\partial \overline{p}_A}\left(\lambda p_L + (1-\lambda)(\gamma \overline{p}_A + (1-\gamma)\overline{p}_B)\right)\right) \\ & \quad > \left(1-\hat{g}(\overline{p}_A;\alpha_A)\right)\left(1 + \frac{\partial \hat{g}(\overline{p}_B;\alpha_B)}{\partial \overline{p}_B}\left(\lambda p_L + (1-\lambda)(\gamma \overline{p}_B + (1-\gamma)\overline{p}_A)\right)\right) \end{aligned}$$

We know that $(1-\hat{g}(\overline{p}_B;\alpha_B)) > (1-\hat{g}(\overline{p}_A;\alpha_A))$ and so it remains to show that

$$\frac{\partial \hat{g}(\overline{p}_A;\alpha_A)}{\partial \overline{p}_A}\left(\lambda p_L + (1-\lambda)(\gamma \overline{p}_A + (1-\gamma)\overline{p}_B)\right) > \frac{\partial \hat{g}(\overline{p}_B;\alpha_B)}{\partial \overline{p}_B}\left(\lambda p_L + (1-\lambda)(\gamma \overline{p}_B + (1-\gamma)\overline{p}_A)\right)$$

which is guaranteed for $\gamma \geq \frac{1}{2}$ and a concave $\hat{g}(\overline{p}_i;\alpha_i)$. Thus, we have shown that $\frac{\partial \overline{p}_A}{\partial p_L} < \frac{\partial \overline{p}_B}{\partial p_L}$. Next, we show that $\frac{\partial \overline{p}_B}{\partial \gamma} > -\frac{\partial \overline{p}_A}{\partial \gamma}$. As before, it suffices to compare $-|J_A(\gamma)|$ and $|J_B(\gamma)|$.

$$\begin{aligned}
-|J_A(\gamma)| &= (1-\hat{g}(\overline{p}_A;\alpha_A))(1-\lambda)(\overline{p}_B - \overline{p}_A)\left(X_B - (1-\lambda)(1-\gamma)(1-\hat{g}(\overline{p}_B;\alpha_B))\right) \\
|J_B(\gamma)| &= (1-\hat{g}(\overline{p}_B;\alpha_B))(1-\lambda)(\overline{p}_B - \overline{p}_A)\left(X_A - (1-\lambda)(1-\gamma)(1-\hat{g}(\overline{p}_A;\alpha_A))\right).
\end{aligned}$$

This reduces to exactly the same condition as before and so we have established that $\frac{\partial \overline{p}_B}{\partial \gamma} > -\frac{\partial \overline{p}_A}{\partial \gamma}$. Thus whenever $\frac{\partial Y_A}{\partial \overline{p}_A}$ is either larger or not much smaller than $\frac{\partial Y_B}{\partial \overline{p}_B}$, Part 1 is negative. A sufficient condition is $\frac{\partial^3 \pi(\overline{p}, \alpha \hat{f}(\overline{p};\alpha))}{\partial \overline{p}^3} \leq 0$.

Part 2 We write

$$Y_A \frac{\partial^2 \overline{p}_A}{\partial p_L \partial \gamma} + Y_B \frac{\partial^2 \overline{p}_B}{\partial p_L \partial \gamma} = Y_A \left\{ \frac{\partial^2 \overline{p}_A}{\partial p_L \partial \gamma} - \frac{\partial \overline{p}_A}{\partial p_L}\left(\frac{\partial \overline{p}_B}{\partial p_L}\right)^{-1} \frac{\partial^2 \overline{p}_B}{\partial p_L \partial \gamma} \right\}. \tag{4}$$

We know from the first order condition that Y_A and Y_B have opposite signs. As Y_i is decreasing in $\overline{p}_i$ and $\overline{p}_A < \overline{p}_B$, it has to be the case that Y_A is positive and Y_B is negative. Thus, we focus on

$$sign\left(\frac{\partial^2 \overline{p}_A}{\partial p_L \partial \gamma} - \frac{\partial \overline{p}_A}{\partial p_L}\left(\frac{\partial \overline{p}_B}{\partial p_L}\right)^{-1} \frac{\partial^2 \overline{p}_B}{\partial p_L \partial \gamma}\right) = sign\left(\frac{\partial \overline{p}_B}{\partial p_L}\frac{\partial^2 \overline{p}_A}{\partial p_L \partial \gamma} - \frac{\partial \overline{p}_A}{\partial p_L}\frac{\partial^2 \overline{p}_B}{\partial p_L \partial \gamma}\right)$$

as $\frac{\partial \overline{p}_B}{\partial p_L} > 0$.

We know that $\frac{\partial \overline{p}_A}{\partial p_L} = \frac{-|J_A(p_L)|}{|J|}$, which leads to

$$\frac{\partial^2 \overline{p}_A}{\partial p_L \partial \gamma} = \frac{|J|\frac{\partial |J_A(p_L)|}{\partial \gamma} - |J_A(p_L)|\frac{\partial |J|}{\partial \gamma}}{|J|^2}$$

We can then write

$$
\begin{aligned}
&\frac{\partial \overline{p}_B}{\partial p_L}\frac{\partial^2 \overline{p}_A}{\partial p_L \partial \gamma} - \frac{\partial \overline{p}_A}{\partial p_L}\frac{\partial^2 \overline{p}_B}{\partial p_L \partial \gamma} \\
=&\frac{|J_B(p_L)|}{|J|}\frac{|J|\frac{\partial |J_A(p_L)|}{\partial \gamma} - |J_A(p_L)|\frac{\partial |J|}{\partial \gamma}}{|J|^2} - \frac{|J_A(p_L)|}{|J|}\frac{|J|\frac{\partial |J_B(p_L)|}{\partial \gamma} - |J_B(p_L)|\frac{\partial |J|}{\partial \gamma}}{|J|^2} < 0 \\
\Leftrightarrow&|J_B(p_L)|\frac{\partial |J_A(p_L)|}{\partial \gamma} - |J_A(p_L)|\frac{\partial |J_B(p_L)|}{\partial \gamma} < 0
\end{aligned} \tag{5}
$$

We have shown that $|J_B(p_L)| > |J_A(p_L)| > 0$. We thus focus on:

$$
\begin{aligned}
\frac{\partial |J_A(p_L)|}{\partial \gamma} =&\frac{\partial \hat{g}(\overline{p}_B;\alpha_B)}{\partial \overline{p}_B}\lambda(1-\lambda)\left(1-\hat{g}(\overline{p}_A;\alpha_A)\right)\left(\overline{p}_B-\overline{p}_A\right) \\
&-2\lambda(1-\lambda)\left(1-\hat{g}(\overline{p}_A;\alpha_A)\right)\left(1-\hat{g}(\overline{p}_B;\alpha_B)\right) < 0 \\
\frac{\partial |J_B(p_L)|}{\partial \gamma} =&\frac{\partial \hat{g}(\overline{p}_A;\alpha_A)}{\partial \overline{p}_A}\lambda(1-\lambda)\left(1-\hat{g}(\overline{p}_B;\alpha_B)\right)\left(\overline{p}_A-\overline{p}_B\right) \\
&-2\lambda(1-\lambda)\left(1-\hat{g}(\overline{p}_A;\alpha_A)\right)\left(1-\hat{g}(\overline{p}_B;\alpha_B)\right).
\end{aligned}
$$

Equation (5) is true whenever $\frac{\partial |J_B(p_L)|}{\partial \gamma} \geq 0$. If $\frac{\partial |J_B(p_L)|}{\partial \gamma} < 0$, then it remains true if $\frac{\partial |J_A(p_L)|}{\partial \gamma} < \frac{\partial |J_B(p_L)|}{\partial \gamma}$ which is equivalent to

$$
\frac{\partial \hat{g}(\overline{p}_B;\alpha_B)}{\partial \overline{p}_B}\lambda(1-\lambda)\left(1-\hat{g}(\overline{p}_A;\alpha_A)\right)\left(\overline{p}_B-\overline{p}_A\right) < \frac{\partial \hat{g}(\overline{p}_A;\alpha_A)}{\partial \overline{p}_A}\lambda(1-\lambda)\left(1-\hat{g}(\overline{p}_B;\alpha_B)\right)\left(\overline{p}_A-\overline{p}_B\right),
$$

which we already established above.

References

Ballester, C., A. Calvó-Armengol, and Y. Zenou (2006). Who's Who in Networks. Wanted: The Key Player. *Econometrica 74*(5), 1403–1417.

Fudenberg, D. and J. Tirole (1991). *Game Theory*. The MIT Press.