MEDIUM AND LONG-RUN PROPERTIES OF LINGUISTIC COMMUNITY EVOLUTION

MICHAEL J. FOX, GEORGIOS PILIOURAS, & JEFF S. SHAMMA

School of Electrical and Computer Engineering, Georgia Institute of Technology
777 Atlantic Drive NW, Atlanta, GA 30332-0250 USA
{*mfox, georgios, shamma* }*@gatech.edu*

We analyze a novel model of the co-evolution of linguistic community structure and language. Intuitively, agents want to communicate well with others in their linguistic community, and similarly, linguistic communities consist of those agents that can communicate effectively amongst themselves. Absent the effects of comunity structure, the model suffers from poor efficiency in the medium-run. Over the long-run, efficiency is attained, but diversity vanishes. When effects of homophily are added to the model we find a more nuanced picture. When the population size is large relative to frequency of stochastic shocks then both diversity and efficiency are observed in the medium-run, but diversity does not survive in the long-run. If stochasic shocks are more frequent then diversity and efficiency can survive the long-run.

1. Introduction

The stunning diversity of the modern language lanscape provokes curiosity as to its origins and persistence. Models of language evolution have tended to consider in isolation mechanisms accounting for either efficiency or diversity. Nominally, a model of the former phenomenon attempts to find conditions leading to a population state where a single, sensible language predominates— a circumstance that is clearly at odds with the latter. Indeed, a differential equation model of language competition has been suggested as evidence that linguistic diversity is a transient phenomenon destined to die out (Abrams & Strogatz, 2003).

Exogenous factors such as geographic isolation (Patriarca & Leppnen, 2004) and language's role as an in-group marker (Dunbar, 1998) have been suggested to account for this discrepancy. More recently, researchers have proposed homophily, the tendency to associate with similar others, as a mechanism to account for the persistence of diverse linguistic communities (Quillinan, Kirby, & Smith, 2010). However, the model views languages as abstract feature vectors. In this framework, the desirability of a language is based soley on similarity with neighbors. In contrast, game-theoretic models of language evolution (Trapa & Nowak, 2000), (Nowak, Plotkin, & Krakauer, 1999) have represented languages explicitly as signalling systems (Lewis, 1969). These systems' symbols and meanings can possess homonymy and symonymy, so that languages have varying degrees of intrinsic ambiguity. However, linguistic communities are not modelled.

In Section 2 we review the so-called language game. It is known that under replicator dynamics, the population can converge to a distribution of languages that is inefficient (Pawlowitsch, 2008), (Huttegger, 2007). We prove a tight lower bound on this efficiency loss, establishing that a particular example from the literature is representative of the worst-case. While diversity can persist in this model, low levels of efficiency can too. This model considers infinite populations— the mass action approach. Infinite populations approximate the behavior of large, finite populations over short time-horizons (Benaim & Weibull, 2003).

In order to account for the inefficiency, researchers have considered mutation-selection dynamics (a perturbation of the replicator dynamics) (Hofbauer & Huttegger, 2008), or finite-population versions of the model (Pawlowitsch, 2007), (Fox & Shamma, 2011a). It is shown in (Fox & Shamma, 2011a) that replicator-like dynamics converge to efficient states in the sense of stochastic stability (Young, 1993). That is, the system spends almost all its time in efficient language states as a parameter describing the frequency of stochastic shocks is made small. We show that the finite-population version of the game is a potential game, which suggests that similar results should exist for most sensible dynamics. These results indicate that, in the long-run, agents overcome inefficient states, but in doing so press out diversity.

We attempt to explain the observed persistence of diversity by suggesting an augmentation of the model that introduces linguistic community structure to agents' interactions. Our model is inspired by a model of opinion formation (Krause, 1997). In that model each agents' opinion is a real number, and at each time step each agent updates her opinion to be the average of the opinions that differ from her own by at most a fixed threshhold. Agents consider opinions that differ by more than the threshhold to be unreasonable. Similarly, our agents define their linguistic community to be those other agents with whom they can communicate above a certain threshhold. At each time step a randomly selected agent updates to the language within her community that achieves the highest utility within that same community. Intuitively, such a model ought to be friendly to diversity because disparate languages can coexist in different communities.

We are able to perform an exact analysis of this model for a restricted set of parameters. In this case we find that only monomorphic language states survive in the long-run. However, this analysis is relevant only when stochastic shocks are extremely rare. Disruptive events with profound implications for the language landscape, such as the reintroduction of modern hebrew in the 20th century, would seem to betray such assumptions. Simulation results are provided for higher levels of randomizing behavior, which suggest a strong tendency towards the formation of distinct linguistic communities. Linguistic coherence is high within these communities, but not between them. Alternatively, we make recourse to convergence rates. It has been shown in a closely related setting that systems like our own may require time to convergence that is exponential in the population size (Shah & Shin, 2010). However, such systems may linger in metastable states over the medium-run (Nimwegen, Crutchfield, & Mitchell, 1997). We present simulation results that suggest efficient but distinct communities can persist in this manner even when they should be expected to vanish over the long-run. A final simulation study shows that the threshhold parameter defining the linguistic community structure has substantial effects on the relative sizes of the communities observed.

2. The language game

We consider a simple language game, first proposed in a substantially similar form in (Lewis, 1969), and reformulated more recently in (Nowak et al., 1999). An agent's speech strategy is an $m \times n$ binary, row-stochastic[a] matrix and her hearing strategy is an $n \times m$ binary, row-stochatic matrix. The linguistic coherence of a particular speaking strategy P against a particular hearing strategy Q is given by $\mathrm{trace}(PQ)$. To see why, expand out

$$\mathrm{trace}(PQ) = \sum_{k=1}^{m} \sum_{j=1}^{n} P_{kj} Q_{jk}. \tag{1}$$

The outer summation considers each of m rows of P. These rows corresponds to the m objects, with the single one in each row indicating the symbol from among the n available that a speaker using P associates with that object. The inner summation for a fixed k equals one if P maps k to the symbol that Q associates with the object k, and zero otherwise. We will consider this basic model of communication from two different perspectives, taking up the infinite-population setting first.

2.1. *Infinite populations*

Suppose that there is a single population of mass equal to one. Each member of the population must choose both speaking and hearing strategies. We confer any ordering on the set of speaking and hearing matrix pairs, or *languages*. There are $m^n n^m$ languages so that the *population state* $\mathbf{x}$ is an element of the $m^n n^m$-dimensional simplex

$$X = \{\mathbf{x} \in \mathbb{R}^{m^n n^m} : x_i \geq 0, \sum_i x_i = 1\}, \tag{2}$$

[a]A matrix is binary and row stochastic if and only if every row has a single element equal to one and all other elements zero.

where x_i indicates the proportion of agents utilizing the i'th language. The fitness of agents speaking the i'th language, (P_i, Q_i), is

$$f_i(\mathbf{x}) = \mathrm{trace}(P_i(\sum_j x_j Q_j)) + \mathrm{trace}((\sum_j x_j P_j) Q_i). \tag{3}$$

In words, an agent's fitness is her coherence achieved from both speaking and hearing, assuming random matching with the population. We study the replicator dynamics

$$\dot{x}_i = x_i(f_i(x) - \sum_j x_j f_j(x)). \tag{4}$$

A population state **x** is a *Neutrally Stable State* (NSS) if

$$\mathbf{x}' f(\mathbf{x}) \geq \mathbf{y}' f(\mathbf{x}) \quad \forall \mathbf{y} \in X, \tag{5}$$

and if $\mathbf{x}' f(\mathbf{x}) = \mathbf{y}' f(\mathbf{x})$ then $\mathbf{x}' f(\mathbf{y}) \geq \mathbf{y}' f(\mathbf{y})$. It has been shown (Pawlowitsch, 2008), (Huttegger, 2007) that the replicator dynamics need not almost always converge to states maximizing average fitness. Indeed we may converge on NSS with average fitness of four for any value of m or n (Hofbauer & Huttegger, 2008). This fact is particularly unsettling taking into account that the maximum average fitness is $2\min\{m, n\}$. We find that this is the worst-case.

Theorem 1: *If **x** is an NSS then* $\sum_i x_i f_i(\mathbf{x}) \geq 4$ *and the bound is tight for all* $m, n \geq 2$.

An NSS may include multiple languages, prompting researchers to view such states as providing opportunities for language diversification (Pawlotisch, Mertikopoulos, & Ritt, 2011). However, the accompanying possibility of such a large efficiency gap prohibits the model from providing a wholistic account of languages evolving into a diverse landscape.

The mass-action heuristic can be shown to approximate sufficiently large populations over limited time spans (Benaim & Weibull, 2003). In order to understand the long-term behavior of the model we must represent the finite population explicitly.

2.2. *Finite populations*

We consider N agents, each utilizing a particular language. Let $(\mathbf{P}, \mathbf{Q})$ be an vector of N languages, one for each agent. Then the fitness of agent $i \in \{1, ..., N\}$ is

$$f_i(\mathbf{P}, \mathbf{Q}) = \mathrm{trace}(P_i \frac{1}{N-1} \sum_{j \neq i} Q_j) + \mathrm{trace}(\frac{1}{N-1} \sum_{j \neq i} P_j Q_i), \tag{6}$$

analogous to the infinite-population model. The long-run behavior of this game under various dynamics is taken up in (Fox & Shamma, 2011b) and (Fox &

Shamma, 2011a). Each of these dynamics assume that at each time step an agent will with probability $\epsilon > 0$ switch to a different language at random. The states that are observed with positive probability in the limit as ϵ is taken to zero are called *stochastically stable* (Young, 1993). For each dynamic it is shown that the stochastically stable states correspond to monomorphic states maximizing average fitness. We next argue that such outcomes should be expected very generally for this game.

A game is *a potential game* (Monderer & Shapley, 1996) if there exists a function $\Phi : \mathcal{A} \to \mathbb{R}$ (the domain being the set of joint strategies) such that for any player i, any joint strategy $\mathbf{s}$, and any strategy s of player i we have

$$f_i(s, \mathbf{s}_{-i}) - f_i(\mathbf{s}) = \Phi(s, \mathbf{s}_{-i}) - \Phi(\mathbf{s}), \tag{7}$$

where $\mathbf{s}_{-i}$ is the vector of strategiess for players other than i. The implication of this definition is that individual optimizing activity is tantamount to optimization of the potential function Φ.

Theorem 2: *The finite-population language game is a potential game with potential function* $\Phi \equiv \frac{1}{2}\sum_{i=1}^{N} f_i$.

Since the potential function is proportional to average fitness, it is not suprising that stochastic evolutionary dynamics tend to maximize average fitness. An intriguing question is how this model can be augmented to account for diversity, while preserving some measure of efficiency. Towards this end we propose a dynamic model that assumes the formation of linguistic communities.

3. Linguistic communities

At each time t we select an agent uniformly at random. This agents neighborhood

$$h_i(\mathbf{P}[t], \mathbf{Q}[t]) \equiv \{j \neq i : \text{trace}(P_i[t]Q_j[t] + P_j[t]Q_i[t]) > r\} \tag{8}$$
$$\cup \{j \neq i : (P_j[t], Q_j[t]) = (P_i[t], Q_i[t])\}, \tag{9}$$

is precisely the agents with whom she can communicate at a level above some fixed threshold $r \in (0, 2\min\{m, n\})$ along with those sharing her language. If she cannot communicate with anyone above the threshhold she picks a new language at random uniformly. Otherwise, with probability $1 - \epsilon$, she selects the language of an agent within her neighborhood that achieves maximum performance relative to her neighborhood, i.e. from amongst the set

$$\arg \max_{j \in h_i(\mathbf{P}[t], \mathbf{Q}[t])} \sum_{k \in h_i(\mathbf{P}[t], \mathbf{Q}[t])} \text{trace}(P_j[t]Q_k[t]) + \text{trace}(P_k[t]Q_j[t]). \tag{10}$$

Or, with small probability $\epsilon > 0$ she chooses a random language instead. All other agents continue with their previous language and a new agent is selected for revision as above.

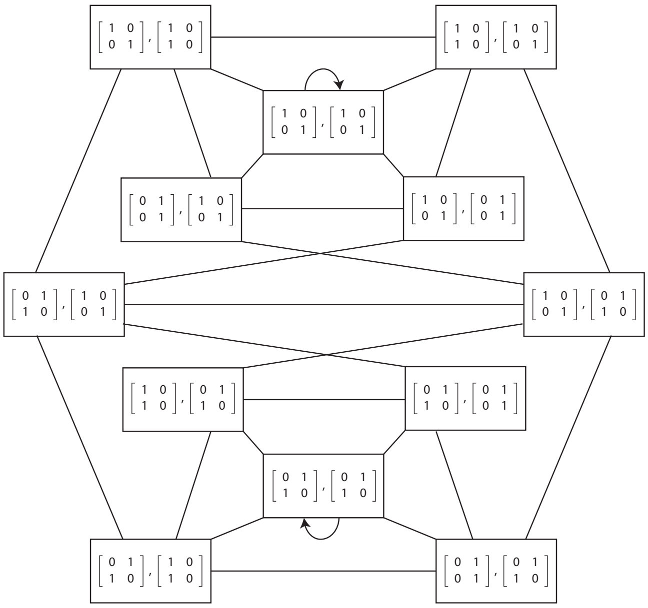

Figure 1. The network structure for $m = n = 2$, $r = 3$.

Different values of m, n, and r will give a range of possible network structures. As an example, we illustrate the structure for $m = n = 2$, $r = 3$ in Fig. 1. An agent's neighbors are the agents who use the languages that her own language is linked to in this graph. Notice that for these parameters only 12 of 16 languages possess links, and only two are linked to themselves.

The model appears contrived because an agent that can communicate effectively within her community will with high probability eschew the opportunity to revise her language in order to communicate effectively with a larger community. We are altogether ignoring the advantages of incumbency that models of language competition concentrate on. It turns out that for small ϵ, even this is not enough.

Theorem 3: *Let $m = n \geq 2$ and $r \in (2(n-1), 2n)$. Then the stochastically stable states of the linguistic community model are the monomorphic states maximizing average fitness.*

Recall that the stochastically stable states are almost all we will see in the long-run. For these parameters, the elaborate linguistic community structure has thus made no difference at all from the viewpoint of stochastic stability . Our analysis does not preclude diversity over the medium-run or for larger values of ϵ, perspectives we now take up.

Stochastic stability characterizes long-run behavior, but such predictions may only become relevant after extraordinary lengths of time. Under more reasonable timescales states that are not stochastically stable may "appear" stable, a phenomenon referred to as metastability. Simulation results illustrated in the leftmost plot of Fig. 2 indicate this can occur for parameter values covered by our theorem. Two languages satisfy $\mathrm{trace}(PQ) = m = n = 2$, these are the *aligned* languages and are represented with dotted lines. The wider dots indicates the more prevalent of the two[b]. The solid line sums over all other languages. Despite eventually settling into monmorphic states, the simulations indicate a metastable epoch where diverse communities thrive.

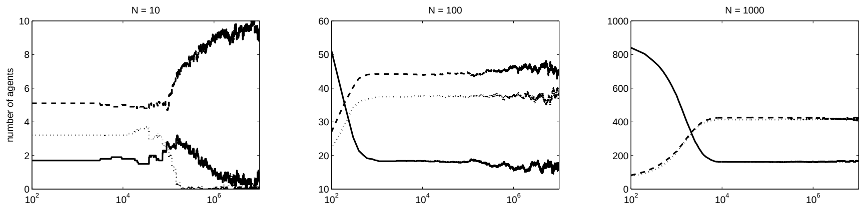

Figure 2. Stability of diversity varies with N; $r = 3, m = n = 2, \epsilon = 10^{-4}$, average of 10 runs.

When ϵ is not small relative to the population size, we observe diverse, efficient linguistic communities even over long time-horizons . The impact of increasing the population size with ϵ fixed is illustrated in Fig. 2. For $\epsilon = 10^{-4}$ we eventually observe monomorphic states for small N, consistent with stochastic stability analysis. As N grows we observe two equally sized internally efficient linguistic communities[c]. Interestingly, the relative community sizes are related to the particular choice of threshhold as illustrated in Fig. 3.

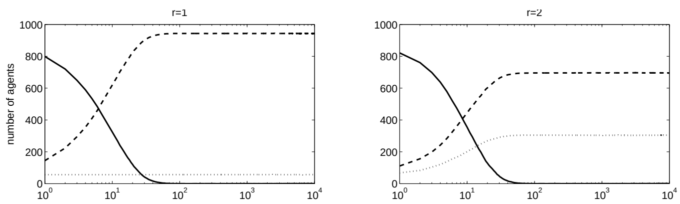

Figure 3. Community sizes vary with r; $m = n = 2, \epsilon = 10^{-4}, N = 1000$, average of 10 runs.

References

Abrams, D. M., & Strogatz, S. H. (2003). Modelling the dynamics of language death. *Nature*, *424*, 900.

Benaim, M., & Weibull, J. W. (2003). Deterministic approximation of stochastic evolution in games. *Econometrica*, *71*(3), 873–903.

[b]The identity of the more prevalent language is allowed to change. Either of the two states that is monomorphic in an aligned language can begin to dominate so, absent our convention, averaging over many runs would give the misleading appearance of diversity for large t

[c]Readers should exercise caution in drawing conclusions from these simulations when N is large because convergence rates generally increase with N. However, even simulations initiated from monomorphic states (not shown) behaved similarly

Dunbar, R. (1998). *Grooming, gossip, and the evolution of language.* Harvard University Press: Cambrdige, MA.

Fox, M., & Shamma, J. (2011a). Language evolution in finite populations. In *50th ieee conference on decision and control.*

Fox, M., & Shamma, J. (2011b). Stochastic stability in language evolution. In *22nd international conference on game theory.*

Hofbauer, J., & Huttegger, S. M. (2008). Feasibility of communication in binary signaling games. *Journal of Theoretical Biology*, *254*(4), 843 - 849.

Huttegger, S. M. (2007). Evolution and the explanation of meaning. *Philosophy of Science*, *74*(1), 1-27.

Krause, U. (1997). Soziale dynamiken mit vielen interakteuren. eine problemskizze. In *Modellierung simul. von dynamiken mit vielen interagierenden akteuren* (p. 37-51).

Lewis, D. (1969). *Convention: A philosophical study.* Harvard Univ. Press, Cambridge, MA.

Monderer, D., & Shapley, L. (1996). Potential games. *Games and Economic Behavior*, *14*, 124–143.

Nimwegen, E. V., Crutchfield, J. P., & Mitchell, M. (1997). Finite populations induce metastability in evolutionary search. *Physics Letters A*, *229*, 144–150.

Nowak, M., Plotkin, J., & Krakauer, D. (1999). The evolutionary language game. *Journal of Theoretical Biology*, *200*(2), 147 - 162.

Patriarca, M., & Leppnen, T. (2004). Modeling language competition. *Physica A: Statistical Mechanics and its Applications*, *338*(1-2), 296 - 299. (Proceedings of the conference A Nonlinear World: the Real World, 2nd International Conference on Frontier Science)

Pawlotisch, C., Mertikopoulos, P., & Ritt, N. (2011). Neutral stability, drift, and the diversification of languages. In *22nd international conference on game theory.*

Pawlowitsch, C. (2007). Finite populations choose an optimal language. *Journal of Theoretical Biology*, *249*(3), 606 - 616.

Pawlowitsch, C. (2008). Why evolution does not always lead to an optimal signaling system. *Games and Economic Behavior*, *63*(1), 203 - 226.

Quillinan, J., Kirby, S., & Smith, K. (2010). Co-evolution of language and social network structure through cultural transmission. In *8th international conference on the evolution of language* (p. 475-476).

Shah, D., & Shin, J. (2010). Dynamics in congestion games. In *Sigmetrics* (p. 107-118).

Trapa, P., & Nowak, M. (2000). Nash equilibria for an evolutionary language game. *Journal of Mathematical Biology*, *41*, 172-188.

Young, H. (1993). The evolution of conventions. *Econometrica*, *61*(1), 57-84.

Appendix A. Appendices

Appendix A.1. *Proof of Theorem 1*

It is straightforward to construct population states satisfying the bound with equality for any m, n. To see this, note that we can associate any population state **x** with the average speaker and hearer,

$$(\bar{P}, \bar{Q}) = (\sum_j x_j P_j, \sum_j x_j Q_j). \tag{11}$$

This way, the set of average languages is simply the product of the set of $m \times n$ row-stochastic matrices and the set of $n \times m$ row-stochastic matrices. In (Hofbauer & Huttegger, 2008), NSS achieving average fitness of four are described for the case of $m = n$. We trivially extend their example to the case of general m, n. Consider $(\bar{P}, \bar{Q})$ given by

$$\left(\begin{bmatrix} \lambda_1 & \lambda_2 & \cdots & \lambda_{n-1} & 0 \\ 0 & 0 & \cdots & 0 & 1 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & \mu_1 & \cdots & \mu_{m-1} \end{bmatrix} \right), \tag{12}$$

where $0 \leq \lambda_i, \mu_j \leq 1$ for all i, j. Clearly $\mathrm{trace}(\bar{P}\bar{Q}) = \sum_i \lambda_i + \sum_j \mu_j = 2$, implying an average fitness of four. That the corresponding population state is an NSS follows from the following lemma, which is merely a combination of Theorem 1 and Lemma 1 in (Pawlowitsch, 2007).

Lemma A1.1: *An average language (P, Q) correponds to an NSS if and only if it satisfies the following four conditions:*

1. $$P \in \underset{\hat{P} \text{ row-stochastic}}{\mathrm{argmax}} \ \mathrm{trace}(\hat{P}Q),$$
2. $$Q \in \underset{\hat{Q} \text{ row-stochastic}}{\mathrm{argmax}} \ \mathrm{trace}(P\hat{Q}),$$
3. *at least one of P, Q has no zero-column, and*
4. *neither P nor Q has a column with multiple maximal elements in $(0, 1)$.*

The first two conditions are neccesary and sufficient for (P, Q) to be a Nash equilibrium and the second two conditions are neccesarry and sufficient for a Nash equilibrium to be an NSS in this setting. In order to prove our theorem, we must show that no NSS can achieve average fitness strictly less than four.

Suppose there exists $(\bar{P}, \bar{Q})$ that is an NSS and satisfies $\mathrm{trace}(\bar{P}\bar{Q}) < 2$, so that it corresponds to an average fitness less than four. It follows that there are $m-1$ rows of $\bar{P}$ that contribute less than one to $\mathrm{trace}(\bar{P}\bar{Q})$. That is, the set

$$\mathcal{R} = \{i : \sum_j \bar{P}_{ij}\bar{Q}_{ji} < 1\}, \tag{13}$$

has cardinality at least $m-1$. The next lemma shows that it suffices to consider $\bar{P}$ for which no two of these rows are the same standard basis vector.

Lemma A1.2: *Suppose $m > 2, n \geq 2$ and (P, Q) is the average language corresponding to an NSS. Further suppose that there exist i_1, i_2 such that $P_{i_1 k} = P_{i_2 k} = 1$ for some k. Then there exists $(\hat{P}, \hat{Q})$ that corresponds to an NSS for the reduced game with dimensions $(m-1), n$ and also satisfies* $\mathrm{trace}(PQ) = \mathrm{trace}(\hat{P}\hat{Q})$.

Proof: Given (P, Q) and assuming without loss of generality that $i_1 = 1$ and $i_2 = 2$ we can construct $(\hat{P}, \hat{Q})$ as

$$\hat{P}_{ij} = \begin{cases} P_{1j}, & i = 1 \\ P_{(i+1)j}, & i \neq 1 \end{cases}, \tag{14}$$

so that $\hat{P}$ consolidates the identical first two rows of P. For Q, we consolidate by combining the first two columns (i_1 and i_2), so that

$$\hat{Q}_{ij} = \begin{cases} Q_{i1} + Q_{i2}, & j = 1 \\ Q_{i(j+1)}, & j \neq 1 \end{cases}. \tag{15}$$

In order to complete the proof of the lemma we must verify both that $(\hat{P}, \hat{Q})$ corresponds to an NSS of the reduced-order game and that $\mathrm{trace}(PQ) = \mathrm{trace}(\hat{P}\hat{Q})$. The latter is easily verified by expanding out,

$$\mathrm{trace}(\hat{P}\hat{Q}) = \sum_i \sum_j \hat{P}_{ij}\hat{Q}_{ji} = \sum_{i\neq 1} \sum_j P_{(i+1)j} Q_{j(i+1)} + \sum_j \hat{P}_{1j}\hat{Q}_{j1} \tag{16}$$

$$= \sum_{i\neq 1} \sum_j P_{(i+1)j} Q_{j(i+1)} + \sum_j \hat{P}_{1j}(Q_{j1} + Q_{j2}) \tag{17}$$

$$= \sum_{i\neq 1} \sum_j P_{(i+1)j} Q_{j(i+1)} + \sum_j P_{1j}Q_{j1} + \sum_j P_{2j}Q_{j2} \tag{18}$$

$$= \mathrm{trace}(PQ), \tag{19}$$

as required.

To verify the NSS property we must show that the four conditions in Lemma A1.1 are satisfied. For the first condition, assume the contrary, i.e. there exists $\tilde{P}$

such that $\mathrm{trace}(\tilde{P}\hat{Q}) > \mathrm{trace}(\hat{P}\hat{Q})$. The matrix $\tilde{P}$ can be expanded into a speaker matrix for the $m \times n$ game given by

$$P^*_{ij} = \begin{cases} \tilde{P}_{1j}, & i \in \{1,2\} \\ \tilde{P}_{(i-1)j}, & i \notin \{1,2\} \end{cases}. \tag{20}$$

We show that the existence of P^* contradicts the supposition that (P, Q) is an NSS. This follows from computing

$$\begin{aligned} \mathrm{trace}(P^*Q) &= \sum_i \sum_j P^*_{ij} Q_{ji} && (21) \\ &= \sum_{i \notin \{1,2\}} \sum_j \tilde{P}_{(i-1)j} Q_{ji} + \sum_j \tilde{P}_{1j} Q_{j1} + \sum_j \tilde{P}_{1j} Q_{j2} && (22) \\ &= \sum_{i \notin \{1,2\}} \sum_j \tilde{P}_{(i-1)j} \hat{Q}_{j(i-1)} + \sum_j \tilde{P}_{1j} \hat{Q}_{j1} && (23) \\ &= \mathrm{trace}(\tilde{P}\hat{Q}) > \mathrm{trace}(\hat{P}\hat{Q}) = \mathrm{trace}(PQ), && (24) \end{aligned}$$

implying that (P, Q) does not satisfy the first condition of Lemma A1.1.

The second condition can be verified similarly. Assume the contrary, i.e. there exists $\tilde{Q}$ such that $\mathrm{trace}(\hat{P}\tilde{Q}) > \mathrm{trace}(\hat{P}\hat{Q})$. We define

$$Q^*_{ij} = \begin{cases} \frac{\tilde{Q}_{i1}}{2}, & j \in \{1,2\} \\ \tilde{Q}_{i(j-1)}, & j \notin \{1,2\} \end{cases}, \tag{25}$$

so that we have

$$\begin{aligned} \mathrm{trace}(PQ^*) &= \sum_i \sum_j P_{ij} Q^*_{ji} && (26) \\ &= \sum_{i \notin \{1,2\}} \sum_i P_{ij} \tilde{Q}_{j(i-1)} + \sum_j P_{1j} \frac{\tilde{Q}_{j1}}{2} + \sum_j P_{2j} \frac{\tilde{Q}_{j1}}{2} && (27) \\ &= \sum_{i \notin \{1,2\}} \sum_j \hat{P}_{(i-1)j} \tilde{Q}_{j(i-1)} + \sum_j \hat{P}_{1j} \tilde{Q}_{j1} && (28) \\ &= \mathrm{trace}(\hat{P}\tilde{Q}) > \mathrm{trace}(\hat{P}\hat{Q}) = \mathrm{trace}(PQ), && (29) \end{aligned}$$

implying that (P, Q) does not satisfy the second condition of Lemma A1.1.

For the third condition of Lemma A1.1 first suppose P has no zero column. Consider the sum of the j'th column of $\hat{P}$. If $j = k$ then

$$\sum_i \hat{P}_{ik} = P_{1k} + \sum_{i \neq 1} P_{(i+1)k} = 1 + \sum_{i \neq 1} P_{(i+1)k} > 0, \tag{30}$$

as required. If $j \neq k$ then

$$\sum_i \hat{P}_{ij} = P_{1j} + \sum_{i \neq 1} P_{(i+1)j} = P_{1j} + P_{2j} + \sum_{i \neq 1} P_{(i+1)j} = \sum_i P_{ij} > 0, \quad (31)$$

as required. Instead suppose Q has no zero column. Aside from the first column, all the columns of $\hat{Q}$ match columns of Q and are hence non-zero. The first column sum is given by

$$\sum_i \hat{Q}_{i1} = \sum_i Q_{i1} + Q_{i2} > \sum_i Q_{i1} > 0, \quad (32)$$

as required.

Finally, we verify the fourth condition. First consider $\hat{P}$. The k'th column has

$$\max_i \hat{P}_{ik} = P_{1k} = 1, \quad (33)$$

by assumption. For any other column $j \neq k$ we have two cases. If $\max_i P_{ij} = 1$ then so does $\max_i \hat{P}_{ij}$. Alternatively if $\max_i P_{ij} \in (0, 1)$ then the cardinality of $\operatorname{argmax}_i \hat{P}_{ij}$ is still one because P_{1j}, the deleted element, is equal to zero. Lastly, consider the first column of $\hat{Q}$. The other columns are unchanged from Q. Suppose that $\max_i \hat{Q}_{i1} \in (0, 1)$. This requires that $\max_i Q_{i1} \in [0, 1)$ and $\max_i Q_{i2} \in [0, 1)$, with at most one of the quantities equal to zero. Assume without loss of generality that $\max_i Q_{i1} \in (0, 1)$. If $\max_i Q_{i2} = 0$ then $\hat{Q}_{i1} = Q_{i1}$ for all i and the condition is satisfied. Therefore, suppose $\max_i Q_{i2} \in (0, 1)$ as well. We claim that

$$\operatorname*{argmax}_i \hat{Q}_{i1} = \operatorname*{argmax}_i Q_{i1} = k = \operatorname*{argmax}_i Q_{i2}, \quad (34)$$

which clearly implies the fourth condition. The first equality follows from the second and third equalities and the definition of $\hat{Q}$. The second and third equalities can be verified by supposing $\operatorname{argmax}_i Q_{i\hat{j}} = \hat{k} \neq k$ for some $\hat{j} \in \{1, 2\}$. We could then define

$$\tilde{P}_{ij} = \begin{cases} P_{ij}, & i \neq \hat{j} \\ 1, & i = \hat{j}, j = \hat{k} \\ 0, & i = \hat{j}, j \neq \hat{k} \end{cases}, \quad (35)$$

so that

$$\operatorname{trace}(\tilde{P}Q) - \operatorname{trace}(PQ) = \sum_j \tilde{P}_{\hat{j}j} Q_{j\hat{j}} - \sum_j P_{\hat{j}j} Q_{j\hat{j}} \quad (36)$$

$$= Q_{\hat{k}\hat{j}} - Q_{k\hat{j}} > 0, \quad (37)$$

which contradicts condition one of Lemma A1.1 for (P, Q). ■

As mentioned above, Lemma A1.2 allows us to assume that $\bar{P}$ contains no two rows that are the same standard basis vector. This is because we can apply Lemma A1.2 inductively until there are no more repeated standard basis vectors. If a counterexample to the theorem exists for the higher-dimensional game with the repeated basis vectors, then the existence of a counter-example is also implied for the lower-dimensional game sans the repeated basis vector. That is, so long as $m > 2$ and $n \geq 2$. It turns out that there are no such NSS when $m = 2$.

Lemma A1.3: *Suppose $m = 2, n \geq 2$ and P's rows are both the same standard basis vector. Then (P, Q) is not the average language corresponding to an NSS.*

Proof: Asssume that (P, Q) is an NSS. We can assume without loss of generality that

$$P = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \end{bmatrix}. \tag{38}$$

It follows that $\mathrm{trace}(PQ) = Q_{11} + Q_{12}$. We claim that

$$1 \in \underset{i}{\mathrm{argmax}}\, Q_{ij}, \text{ for } j \in \{1, 2\}. \tag{39}$$

To see this, assume the contrary, i.e. that there exists $\hat{i} \neq 1, \hat{j} \in \{1, 2\}$ such that $Q_{\hat{i}\hat{j}} > Q_{1\hat{j}}$. Then construct

$$\tilde{P}_{ij} = \begin{cases} P_{ij}, & i \neq \hat{j} \\ 1, & i = \hat{j}, j = \hat{i} \\ 0, & i = \hat{j}, j \neq \hat{i} \end{cases}, \tag{40}$$

and observe that

$$\mathrm{trace}(\tilde{P}Q) - \mathrm{trace}(PQ) = Q_{\hat{i}\hat{j}} - Q_{1\hat{j}} > 0, \tag{41}$$

which contradicts our assumption that (P, Q) is an NSS because the first condition of Lemma A1.1 is violated. Next, assume without loss of generality that $1 > Q_{11} \geq \frac{1}{2}$, implying $0 < Q_{12} \leq \frac{1}{2}$. The strict inequalities are implied by the third condition of Lemma A1.1 along with the preceding claim. The second row must also sum to one, so its maximum element must be at least $\frac{1}{2}$. If the maximum element is in the first column then it must be strictly less than Q_{11} due to the fourth condition of Lemma A1.1. It follows that

$$Q_{22} = 1 - Q_{21} > 1 - Q_{11} = Q_{12}, \tag{42}$$

which contradicts our claim. If the maximum element of the second row is in the second column then $Q_{22} \geq \frac{1}{2} \geq Q_{12}$, which contradicts our claim since the fourth condition of Lemma A1.1 implies $Q_{22} \neq Q_{12}$. ■.

The two preceding lemmas allow us to assume that no two rows in $\mathcal{R}$ are the same standard basis vector. It follows that there exists $\hat{i} \in \mathcal{R}$ and $\hat{j}$ such that

$$\{\hat{i}\} = \operatorname*{argmax}_i \bar{P}_{i\hat{j}}. \tag{43}$$

In other words, some element in one of the rows in $\mathcal{R}$ is the unique maximum element in its column. To see this, assume the contrary. That is, each column of P either has its unique maximum element in the row not included in $\mathcal{R}$, or has multiple maximum elements. If a column has multiple maximum elements then those elements must all be equal to one by the fourth condition of Lemma A1.1. Since no two rows of $\bar{P}$ are the same standard basis vector this implies that the row not in $\mathcal{R}$ contains the other one, with its other elements being zero. Remaining columns cannot have multiple maximum elements or their maximum elements in the row not in $\mathcal{R}$, a contradiction. If all column maxima are unique, then one of the rows in $\mathcal{R}$ must contain such a maximum by row-stochasticity.

The existence of $\hat{i}, \hat{j}$ implies that $\bar{Q}_{\hat{j}\hat{i}} = 1$. If it did not then we could construct

$$\tilde{Q}_{ij} = \begin{cases} \bar{Q}_{ij}, & i \neq \hat{j} \\ 1, & i = \hat{j}, j = \hat{i} \\ 0, & i = \hat{j}, j \neq \hat{i} \end{cases}, \tag{44}$$

so that

$$\operatorname{trace}(\bar{P}\tilde{Q}) - \operatorname{trace}(\bar{P}\bar{Q}) = 1 - \bar{Q}_{\hat{j}\hat{i}} > 0, \tag{45}$$

which contradicts the second condition of Lemma A1.1. We conclude the proof of Theorem 1 by demonstrating that $\hat{i} \notin \mathcal{R}$, a contradiction. Otherwise we could construct

$$\tilde{P}_{ij} = \begin{cases} \bar{P}_{ij}, & i \neq \hat{i} \\ 1, & i = \hat{i}, j = \hat{i} \\ 0, & i = \hat{i}, j \neq \hat{j} \end{cases}, \tag{46}$$

so that

$$\operatorname{trace}(\tilde{P}\bar{Q}) - \operatorname{trace}(\bar{P}\bar{Q}) = 1 - \sum_j \bar{P}_{\hat{i}j}\bar{Q}_{j\hat{i}} > 0, \tag{47}$$

which contradcits the second condition of Lemma A1.1, where the inequality is simply the definition of membership in $\mathcal{R}$. ∎

Appendix A.2. ***Proof of Theorem 2***

Let $(\mathbf{P}, \mathbf{Q})$ and $(\hat{\mathbf{P}}, \hat{\mathbf{Q}})$ differ only in the language of player $\hat{\imath}$. Then we have $\Phi(\hat{\mathbf{P}}, \hat{\mathbf{Q}}) - \Phi(\mathbf{P}, \mathbf{Q})$

$$= \frac{1}{2}\sum_i f_i(\hat{\mathbf{P}}, \hat{\mathbf{Q}}) - \frac{1}{2}\sum_i f_i(\mathbf{P}, \mathbf{Q}) \tag{48}$$

$$= \frac{1}{2}\sum_i \left[\mathrm{trace}(\hat{P}_i \frac{1}{N-1}\sum_{j\neq i}\hat{Q}_j) + \mathrm{trace}(\frac{1}{N-1}\sum_{j\neq i}\hat{P}_j\hat{Q}_i)\right] \tag{49}$$

$$- \frac{1}{2}\sum_i \left[\mathrm{trace}(P_i \frac{1}{N-1}\sum_{j\neq i}Q_j) + \mathrm{trace}(\frac{1}{N-1}\sum_{j\neq i}P_jQ_i)\right] \tag{50}$$

$$= \frac{1}{2(N-1)}\sum_{i\neq\hat{\imath}}\sum_{j\notin\{i,\hat{\imath}\}} \mathrm{trace}(P_iQ_j + P_jQ_i) \tag{51}$$

$$+ \frac{1}{N-1}\sum_{j\neq\hat{\imath}}\mathrm{trace}(\hat{P}_{\hat{\imath}}\hat{Q}_j + \hat{P}_j\hat{Q}_{\hat{\imath}}) \tag{52}$$

$$- \frac{1}{2(N-1)}\sum_{i\neq\hat{\imath}}\sum_{j\notin\{i,\hat{\imath}\}} \mathrm{trace}(P_iQ_j + P_jQ_i) \tag{53}$$

$$- \frac{1}{N-1}\sum_{j\neq\hat{\imath}}\mathrm{trace}(P_{\hat{\imath}}Q_j + P_jQ_{\hat{\imath}}) \tag{54}$$

$$= \mathrm{trace}(\hat{P}_{\hat{\imath}} \frac{1}{N-1}\sum_{j\neq\hat{\imath}}\hat{Q}_j) + \mathrm{trace}(\frac{1}{N-1}\sum_{j\neq\hat{\imath}}\hat{P}_j\hat{Q}_{\hat{\imath}}) \tag{55}$$

$$- \mathrm{trace}(P_{\hat{\imath}} \frac{1}{N-1}\sum_{j\neq\hat{\imath}}Q_j) + \mathrm{trace}(\frac{1}{N-1}\sum_{j\neq\hat{\imath}}P_jQ_{\hat{\imath}}) \tag{56}$$

$$= f_{\hat{\imath}}(\hat{\mathbf{P}}, \hat{\mathbf{Q}}) - f_{\hat{\imath}}(\mathbf{P}, \mathbf{Q}), \tag{57}$$

as required. ■

Appendix A.3. ***Proof of Theorem 3***

The proof utilizes the resistance tree method (Young, 1993), which we here review only briefly. Resistance trees are a tool used to characterize the stochastically stable states of Markov chains satisfying certain technical conditions. Let $\mathcal{P}_n^\epsilon$ be the family of Markov chains (in the parameter ϵ) described in our linguistic community model for any fixed value of $m = n \geq 2$. We assume $r = 2n - 1$, noting that the proof is identical for any other value in the specified range. The set of states of $\mathcal{P}^\epsilon$, Z, is the set of all possible assignments of the n^{2n} languages to each of the N players. A state $z \in Z$ is **stochastically stable** if

$$\lim_{\epsilon \to 0} \mu^{\epsilon}(z) > 0, \tag{58}$$

where μ^{ϵ} is the unique stationary distribution of the Markov chain with parameter ϵ. It is straightforward to verify that $\mathcal{P}^{\epsilon}$ is a regular perturbed Markov process allowing us to utilize the resistance tree method to identify its stochastically stable states. Let $Z_1, ..., Z_J$ denote the recurrent communication classes of $\mathcal{P}^0$, the Markov chain obtained by substituting $\epsilon = 0$ into $\mathcal{P}^{\epsilon}$. A state z is stochastically stable if and only if its entire recurrent communication class is, a fact which allows us to characterize stochastic stability at the level of recurrent communication classes as oppose to individual states. Recall that at each time step, players randomize with probability ϵ. The **resistance**[d] between two recurrent communication classes Z_i and Z_j is the minimum number of such randomization events required to transition between the two classes, denoted by $r(Z_i, Z_j)$.

Consider a graph G whose vertex set is the set of recurrent communication classes. A Z_i-tree T is a spanning tree in G such that for any vertex $Z_j, j \neq i$ there is a unique directed path from Z_j to Z_i. We define

$$\gamma(Z_i) = \min_{T \in \mathcal{T}_{Z_i}} \sum_{(Z_j, Z_k) \in T} r(Z_j, Z_k), \tag{59}$$

where $\mathcal{T}_{Z_i}$ is the set of all Z_i trees in G, which we refer to as the **stochastic potential** of Z_i. The stochastically stable recurrent communication classes are precisely those having minimum stochastic potential.

Before proceeding, we briefly outline our proof strategy. Recall that Theorem 3 claims that the stochastically stable states are all monomorphic states in aligned languages. That is, they are singleton recurrent communication classes, or absorbing states, in which every player utilizes some language (P, Q) satisfying $\mathrm{trace}(PQ) = 2n$— the maximum attainable. Let $\mathcal{O}$ refer to the set of states satisfying the above conditions. Given a recurrent class Z_k we can trivially lower bound its stochastic potential

$$\gamma(Z_k) \geq \mu(Z_k) \equiv \sum_{i \neq k} \min_{j \neq i} r(Z_i, Z_j). \tag{60}$$

Since all resistance trees for a given recurrent class have the same set of source nodes, we know that, at best, the minimum resistance tree achieves the resistance between each of these sources and its minimum resistance destination. This bound is not always tight because the "greedy" graphs we implicitly construct need not satisfy the connectivity requirements. We will show that for any $x, y \in \mathcal{O}$ we have

$$\gamma(y) = \gamma(x) = \mu(x) \equiv \gamma_{\mathcal{O}}. \tag{61}$$

[d]Note that this is not the general definition of resistance, but agrees with it for our process.

We then establish stochastic stability of $\mathcal{O}$ by showing that for any $x \notin \mathcal{O}$ we have $\gamma(x) \geq \mu(x) > \gamma_{\mathcal{O}}$.

From here on we refer to recurrent communication classes as just recurrent classes for brevity. Not all recurrent classes are absorbing states. Our proof will not characterize the recurrent classes aside from a few key features. We first note that only languages having at most one total zero column among their two matrices can appear in recurrent states because if a language (P, Q) has two or more total zero columns then

$$\mathrm{trace}(\hat{P}Q + P\hat{Q}) \leq 2n - 2 \quad \forall(\hat{P}, \hat{Q}). \tag{62}$$

Players speaking these languages always randomize when given revision oppurtunities because they cannot communicate with anyone above the threshhold.

The unperturbed process ($\epsilon = 0$) is not innovative (i.e. new languages never appear), so the set of languages in each state in a recurrent class is the same. More formally, consider any two states $z = (\mathbf{P}, \mathbf{Q})$ and $\hat{z} = (\hat{\mathbf{P}}, \hat{\mathbf{Q}})$ contained in a single recurrent class Z_i. For any j there must exist k such that $(\hat{P}_k, \hat{Q}_k) = (P_j, Q_j)$. While each state in a recurrent class must have the same support over the set of all languages, the actual number of players speaking each language can vary from state to state. However, the number of players speaking each aligned language is constant across all of the states in the recurrent class. This is because agents speaking aligned languages never change languages so long as they have neighbors, a claim we establish with the following lemma.

Lemma A3.1: *Suppose that the state* $(\boldsymbol{P}, \boldsymbol{Q})$ *contains an aligned language* (P_i, Q_i) *with* $|h_i(\boldsymbol{P}, \boldsymbol{Q})| \geq 1$ *then each*

$$\hat{i} \in \arg \max_{j \in h_i(\boldsymbol{P}, \boldsymbol{Q})} \sum_{k \in h_i(\boldsymbol{P}, \boldsymbol{Q})} \mathrm{trace}(P_j Q_k + P_k Q_j), \tag{63}$$

satisfies $(P_{\hat{i}}, Q_{\hat{i}}) = (P_i, Q_i)$.

Proof: By the definition of the neighborhood $h_i(\mathbf{P}, \mathbf{Q})$ we have

$$\mathrm{trace}(P_i Q_j + P_j Q_i) \geq 2n - 1, \tag{64}$$

for each $j \in h_i(\mathbf{P}, \mathbf{Q})$. Neighbors of an agent speaking an aligned language either speak the same aligned language or possess a zero column. If agent $j \in h_i(\mathbf{P}, \mathbf{Q})$ possess a zero column then for any $k \in h_i(\mathbf{P}, \mathbf{Q})$ we have

$$\mathrm{trace}(P_k Q_j + P_j Q_k) \leq 2n - 1, \tag{65}$$

and $\mathrm{trace}(P_j Q_j + P_j Q_j) \leq 2n - 2$, so (P_i, Q_i) outperforms (P_j, Q_j) against all members of $h_i(\mathbf{P}, \mathbf{Q})$ and does so strictly against (P_j, Q_j). ■

Since the number of agents speaking any aligned language is non-decreasing for any state trajectory in the recurrent class, all states in each recurrent class must have the same number of agents speaking each aligned language.

Consider recurrent classes containing one or more aligned languages. A probability ϵ event is sufficient to move one agent from a language that is not aligned to one of the aligned languages present in the recurrent class. It does not matter which state we apply the perturbation from. This new state may be transient, but we are guaranteed to reach a new recurrent class with stricly more agents speaking aligned languages. This fact follows from the preceding lemma. Proceeding like this we can reach an abosrbing state in which all agents speak aligned languages. Agents can then switch from one aligned language to another via probability ϵ events so that we reach a state in $\mathcal{O}$ and we required only transitions between absorbing states having resistance equal to one[e].

Next, consider a recurrent class Z_k without any aligned languages. In this case each state $(\mathbf{P}, \mathbf{Q}) \in Z_k$ contains some agent i achieving

$$\mathrm{trace}(P_i Q_i) = \max_j \mathrm{trace}(P_j Q_j) \equiv c(Z_k). \tag{66}$$

We can reach a state that increases this constant by one via a single ϵ probability event. Further, this can be done in such a manner that the agent speaking the new language will have a non-empty neighborhood.

Lemma A3.2: *Suppose* $\mathrm{trace}(PQ) \leq n-1$*. Then there exists another language* $(\hat{P}, \hat{Q})$ *satisftying*

$$\mathrm{trace}(\hat{P}\hat{Q}) = \mathrm{trace}(PQ) + 1, \tag{67}$$

and

$$\mathrm{trace}(\hat{P}Q + P\hat{Q}) \geq r = 2n - 1. \tag{68}$$

Proof: Assume that P has no zero columns and let $\hat{Q} = P'$. The proof when P has a zero column is symmetric. Let

$$C = \{j : \sum_k P_{jk} Q_{kj} = 0\}, \tag{69}$$

the set of indices of columns of Q that do not contribute to $\mathrm{trace}(PQ)$. The matrix Q has at most one zero column. First, suppose Q has a zero column and let

$$C_1 = \{j \in C : \sum_k Q_{kj} > 0\}, \tag{70}$$

the set of indices in C of non-zero columns of Q. Next, let

$$\hat{P}_{ij} = \begin{cases} 1, & i \in C_1, j = \min\{\mathrm{argmax}_k\, Q_{ki}\} \\ 0, & i \in C_1, j \neq \min\{\mathrm{argmax}_k\, Q_{ki}\} \\ P_{ij}, & \text{otherwise.} \end{cases} \tag{71}$$

[e]We will ignore redundancy in our construction since we can always merge paths to eliminate any redundancies in a resistance tree at the end

This way, each row of $\hat{P}$ in C_1 is confined to the support of the corresponding column in Q. Thus,

$$\begin{aligned}
\operatorname{trace}(\hat{P}Q) &= \sum_{i\in C_1}\sum_j \hat{P}_{ij}Q_{ji} + \sum_{i\in C^c}\sum_j \hat{P}_{ij}Q_{ji} + \sum_{i\in C-C_1}\sum_j \hat{P}_{ij}Q_{ji} && (72)\\
&\geq \sum_{i\in C_1}\sum_j \hat{P}_{ij}Q_{ji} + \sum_{i\in C^c}\sum_j \hat{P}_{ij}Q_{ji} && (73)\\
&= \sum_{i\in C_1}\sum_{j=\min\{\operatorname{argmax}_k Q_{ki}\}} Q_{ji} + \sum_{i\in C^c}\sum_j \hat{P}_{ij}Q_{ji} && (74)\\
&= |C_1| + \sum_{i\in C^c}\sum_j P_{ij}Q_{ji} = |C_1| + |C^c| = n-1, && (75)
\end{aligned}$$

which combined with $\operatorname{trace}(P\hat{Q}) = \operatorname{trace}(PP') = n$ establishes the first claim of the lemma. The second claim of the lemma is verified by computing

$$\begin{aligned}
\operatorname{trace}(\hat{P}\hat{Q}) &= \sum_{i\in C_1}\sum_j \hat{P}_{ij}\hat{Q}_{ji} + \sum_{i\in C^c}\sum_j \hat{P}_{ij}\hat{Q}_{ji} + \sum_{i\in C-C_1}\sum_j \hat{P}_{ij}\hat{Q}_{ji} && (76)\\
&\geq \sum_{i\in C^c}\sum_j \hat{P}_{ij}\hat{Q}_{ji} + \sum_{i\in C-C_1}\sum_j \hat{P}_{ij}\hat{Q}_{ji} && (77)\\
&= \sum_{i\in C^c}\sum_j P_{ij}P_{ij} + \sum_{i\in C-C_1}\sum_j P_{ij}P_{ij} && (78)\\
&= |C^c| + |C-C_1| = \operatorname{trace}(PQ) + 1. && (79)
\end{aligned}$$

To complete the proof of the lemma we instead suppose that Q has no zero column. We define the set $C_0 \subset C$ so that $j \in C_0 \Rightarrow \sum_k Q_{kj} > 0$ and $|C_0| = |C_1| - 1$. Put another way, C_0 is any set obtained by removing any one column index from C_1. Next, let

$$\hat{P}_{ij} = \begin{cases} Q_{ji}, & i \in C_0 \\ P_{ij}, & \text{otherwise,} \end{cases} \tag{80}$$

giving,

$$\begin{aligned}
\operatorname{trace}(\hat{P}Q) &= \sum_{i\in C_0}\sum_j \hat{P}_{ij}Q_{ji} + \sum_{i\in C^c}\sum_j \hat{P}_{ij}Q_{ji} + \sum_{i\in C-C_0}\sum_j \hat{P}_{ij}Q_{ji} && (81)\\
&\geq \sum_{i\in C_0}\sum_j \hat{P}_{ij}Q_{ji} + \sum_{i\in C^c}\sum_j \hat{P}_{ij}Q_{ji} && (82)\\
&= \sum_{i\in C_0}\sum_j Q_{ji}Q_{ji} + \sum_{i\in C^c}\sum_j \hat{P}_{ij}Q_{ji} && (83)\\
&= |C_0| + \sum_{i\in C^c}\sum_j P_{ij}Q_{ji} = |C_0| + |C^c| = n-1, && (84)
\end{aligned}$$

as required. Lastly,

$$\text{trace}(\hat{P}\hat{Q}) = \sum_{i\in C_0}\sum_j \hat{P}_{ij}\hat{Q}_{ji} + \sum_{i\in C^c}\sum_j \hat{P}_{ij}\hat{Q}_{ji} + \sum_{i\in C-C_0}\sum_j \hat{P}_{ij}\hat{Q}_{ji} \tag{85}$$

$$\geq \sum_{i\in C^c}\sum_j \hat{P}_{ij}\hat{Q}_{ji} + \sum_{i\in C-C_0}\sum_j \hat{P}_{ij}\hat{Q}_{ji} \tag{86}$$

$$= \sum_{i\in C^c}\sum_j P_{ij}P_{ij} + \sum_{i\in C-C_0}\sum_j P_{ij}P_{ij} \tag{87}$$

$$= |C^c| + |C - C_0| = \text{trace}(PQ) + 1, \tag{88}$$

completing the proof. ■

While the new state may be transient, we reach a recurrent class containing the new language. This is because the last speaker of this new language never abandons her language so long as she has a neighbor.

Lemma A3.3: *Suppose that the state* $(\boldsymbol{P}, \boldsymbol{Q})$ *contains a language* (P_i, Q_i) *with* $|h_i(\boldsymbol{P}, \boldsymbol{Q})| \geq 1$ *such that for all* $j \neq i$ *we have* $(P_j, Q_j) \neq (P_i, Q_i)$*. Further suppose*

$$\text{trace}(P_i Q_i) > \max_{j\neq i} \text{trace}(P_j Q_j), \tag{89}$$

then

$$i = \arg\max_{j\in h_i(\boldsymbol{P},\boldsymbol{Q})} \sum_{k\in h_i(\boldsymbol{P},\boldsymbol{Q})} \text{trace}(P_j Q_k + P_k Q_j). \tag{90}$$

Proof: By the definition of the neighborhood $h_i(\mathbf{P}, \mathbf{Q})$ and the uniqueness of (P_i, Q_i) we have $\text{trace}(P_i Q_j + P_j Q_i) \geq 2n - 1$ for each $j \in h_i(\mathbf{P}, \mathbf{Q})$. Each other language $j \in h_i(\mathbf{P}, \mathbf{Q})$ has $\text{trace}(P_j Q_j) < \text{trace}(P_i Q_i) \leq n$ so it achieves a payoff of at most $2n - 2$ against itself, while (P_i, Q_i) achieves at least $2n - 1$. Since (P_i, Q_i) outperforms each language strictly against at least one other language in the neighborhood (namely, the language itself), it need only match that language against all other languages. Thus it is sufficient to show that for any two agents $k, j \in h_i(\mathbf{P}, \mathbf{Q})$ with $k \neq j$ we have $\text{trace}(P_k Q_j + P_j Q_k) \leq 2n - 1$. Assume the contrary, i.e. that there exist two agents k and j with $k \neq j$ satisfying $\text{trace}(P_k Q_j + P_j Q_k) = 2n$. This requires $P_k = Q'_j$ and $P_j = Q'_k$. Now since j is in $h_i(\mathbf{P}, \mathbf{Q})$ we know that either $P_j = Q'_i$ or $Q_j = P'_i$ because one of the trace terms must equal n. We will deal with the former case only since the latter will then follow from symmetry. By the same reasoning we have that either $P_k = Q'_i$ or $Q_k = P'_i$. If $P_k = Q'_i$ then

$$n = \text{trace}(P_j Q_k) = \text{trace}(P_k Q_k) < n, \tag{91}$$

a contradiction. If $Q_k = P'_i$ then

$$n = \text{trace}(P_j Q_k) = \text{trace}(Q'_i P'_i) = \text{trace}(P_i Q_i). \tag{92}$$

If $\mathrm{trace}(P_iQ_i) = n$ then all its neighbors possess a zero column, so that at least one of the requirements $P_k = Q'_j$ or $P_j = Q'_k$ will violate row stochasticity. ■

Of course, we must guarantee that she continues to have a neighbor on the way to the recurrent class. The next lemma establishes that the last of the neighbors of the agent speaking the new language never abandons her language. That is, unless she abandons her language for the new language. This can happen only if the new language is aligned, but in that case we have reached the scenario described above and are done.

Lemma A3.4: *Suppose that the state $(\boldsymbol{P}, \boldsymbol{Q})$ contains a language (P_i, Q_i) with $|h_i(\boldsymbol{P}, \boldsymbol{Q})| \geq 1$ such that for all $j \neq i$ we have $(P_j, Q_j) \neq (P_i, Q_i)$. Further suppose*

$$\mathrm{trace}(P_iQ_i) \leq n - 1, \tag{93}$$

then either (i),

$$i = \arg \max_{j \in h_i(\boldsymbol{P},\boldsymbol{Q})} \sum_{k \in h_i(\boldsymbol{P},\boldsymbol{Q})} \mathrm{trace}(P_jQ_k + P_kQ_j), \tag{94}$$

or (ii), there exists $(\hat{P}, \hat{Q})$ such that for all

$$\hat{i} \in \arg \max_{j \in h_i(\boldsymbol{P},\boldsymbol{Q})} \sum_{k \in h_i(\boldsymbol{P},\boldsymbol{Q})} \mathrm{trace}(P_jQ_k + P_kQ_j), \tag{95}$$

$(P_{\hat{i}}, Q_{\hat{i}}) = (\hat{P}, \hat{Q})$ *and* $\mathrm{trace}(\hat{P}\hat{Q}) = n$.

Proof: We know that for each $j \in h_i(\mathbf{P}, \mathbf{Q})$ we have

$$\mathrm{trace}(P_iQ_j + P_jQ_i) \geq 2n - 1. \tag{96}$$

Consider any two agents $k, j \in h_i(\mathbf{P}, \mathbf{Q})$ and assume

$$\mathrm{trace}(P_kQ_j + P_jQ_k) = 2n. \tag{97}$$

If this is not possible then (i) obtains. Supposing it is possible we have $P_k = Q'_j$ and $P_j = Q'_k$. Now since j is in $h_i(\mathbf{P}, \mathbf{Q})$ we know that either $P_j = Q'_i$ or $Q_j = P'_i$ because one of the trace terms must equal n. We will deal with the former case only since the latter will then follow from symmetry. By the same reasoning we have that either $P_k = Q'_i$ or $Q_k = P'_i$. If $Q_k = P'_i$ then

$$n = \mathrm{trace}(P_jQ_k) = \mathrm{trace}(Q'_iP'_i) = \mathrm{trace}(P_iQ_i), \tag{98}$$

a contradiction. Thus $P_k = Q'_i$ so that

$$n = \mathrm{trace}(P_jQ_k) = \mathrm{trace}(P_kQ_k), \tag{99}$$

so that if (i) does not obtain, then (ii) obtains because only aligned languages can outperform (P_i, Q_i) against its own neighbors. ■

We can apply the above method inductively so that $c(Z_k)$ increases by one for each recurrent class Z_k visited. We eventually reach a recurrent class containing an aligned language, from which point we have already established the existence of a suitable path to a state in $\mathcal{O}$. From these states in $\mathcal{O}$ all departing edges have resistance at least two. For a resistance tree rooted at a state in $x \in \mathcal{O}$, consider $y \in \mathcal{O}, y \neq x$. From any such state y we can move two players to the aligned language in x, giving a new abosrbing state that achieves the minimum resistance from y of two. Then, we can move one player at a time to the aligned language in x, achieving a resistance of one for each abosrbing state on our way to x. It follows that

$$\gamma(x) = \mu(x) = \gamma_{\mathcal{O}} = 2(|\mathcal{O}| - 1) + |\mathcal{O}^c| \tag{100}$$

For any other recurrent class $y \notin \mathcal{O}$ it is sufficient to note that any resistance tree has one more edge emanating from a state in $\mathcal{O}$, so that

$$\gamma(y) \geq \mu(y) = 2|\mathcal{O}| + |\mathcal{O}^c| - 1 > \gamma_{\mathcal{O}}, \tag{101}$$

completing the proof. ■