2011 50th IEEE Conference on Decision and Control and
European Control Conference (CDC-ECC)
Orlando, FL, USA, December 12-15, 2011

Language Evolution in Finite Populations

Michael J. Fox and Jeff S. Shamma

Abstract— **We study a simple game-theoretical model of language evolution in finite populations. This model is of particular interest due to a surprising recent result for the infinite population case: under replicator dynamics, the population game converges to socially inefficient outcomes from a set of initial conditions with non-zero Lesbegue measure. If finite population models do not exhibit this feature then support is lent to the idea that small population sizes are a key ingredient in the emergence of linguistic coherence. It has been argued elsewhere that evolution supports efficient languages in finite populations using the method of comparing fixation probabilities of single mutant invaders to the inverse of the population size. We instead analyze an alternative generalization of replicator dynamics to finite populations that leads to the emergence of linguistic coherence in an absolute sense. After a long enough period of time, linguistic coherence is observed with arbitrarily high probability as a mutation rate parameter is taken to zero. We also discuss several variations on our model.**

I. Introduction

A. Background

It is difficult to discount the import of language in the success of our species. Human language allows us to spread information at speeds that vastly outstrip the pace of biological evolution. Thus language can be seen as the technology that enables evolutionary change on cultural timescales. Nevertheless, how language first emerged remains somewhat of a mystery. Compounding the issue is a scarcity of physical evidence of the earliest speakers [1]. Two novel approaches to the study of language evolution have emerged in recent decades: genomics [2] and mathematical modeling (for a review, see for instance [3]). We concentrate on the latter. Mathematical modeling of language evolution is especially useful for checking the internal consistency of proposed theories. Alternatively, this endeavor is capable of providing insights into language learning in artificially intelligent systems [4], [5].

A popular approach to explaining language origins is the suggestion that the first languages were simple, possibly gestural [1] linkings from object to symbol. These proto-languages are the predecessors of modern compositional languages. The fundamental problem with the emergence of useful proto-languages is that of cooperation [6]. It is advantageous for many members of a population to associate symbols with objects consistently, but how does such a convention emerge? Invoking a particular symbol to refer to an object is only useful after a significant portion of the population has already adopted such a mapping. Game theory has proven to be a useful framework for studying these simple proto-language models [7], [8], [9], [10], [11], [12], [3].

We mention in passing that the problem we study can also be interpreted as a model of economic signaling [7], [8], [12], [13], although we do not explore this possibility here.

B. The Language Game

We consider a simple language game, first proposed in a substantially similar form in [13], and reformulated more recently in [9]. Each player's strategy (or *language*) is a pair of matrices $(P, Q) \in \mathbb{B}^{m\times n} \times \mathbb{B}^{n\times m} \equiv \mathcal{L}_{m\times n}$, where $\mathbb{B}^{m\times n}$ is the set of binary (having elements from $\{0, 1\}$), row-stochastic $m \times n$ matrices and $\mathcal{L}_{m\times n}$ is the set of *languages*. There are n^2m^2 languages in $\mathcal{L}_{m\times n}$. We refer to the two matrices as the *speaker* and *hearer* matrices, respectively. The speaker matrix maps objects to symbols, and the hearer matrix maps symbols to objects. Every player has the same set of languages available to them. The utility of player i with language (P_i, Q_i) is

$$u_i((P_i, Q_i), (\bar{P}, \bar{Q})) \equiv \frac{1}{2}\mathbf{Tr}(P_i\bar{Q}) + \frac{1}{2}\mathbf{Tr}(\bar{P}Q_i)$$

where $(\bar{P}, \bar{Q})$ are the average of the speaker and hearer matrices, respectively, over the entire population. We depart from the more conventional notation of utilities depending on the joint strategies to emphasize that individuals interact with the entire population and do so anonymously. Note that u_i does not depend on i other than through (P_i, Q_i). The two terms on the right hand side correspond to speaking and hearing, respectively. We can rewrite one of these terms as

$$\mathbf{Tr}(PQ) = \sum_{k=1}^{n}\sum_{j=1}^{m} P_{kj}Q_{jk}$$

where P_{kj} is the kj^{th} element of P and similarly for Q. We interpret this as follows: The internal summation is for a fixed object k. Only a single P_{kj} equals one due to the row-stochasticity. This is the symbol j that the speaker matrix P associates with object k. If the hearer Q associates symbol j with object k (i.e. $Q_{jk} = 1$) then there is a contribution of one to the utility for object k. The total utility is computed by summing over the objects and weighting contributions from speaking and hearing equally. We include (P_i, Q_i) in $(\bar{P}, \bar{Q})$ in order to streamline the notation, but all of our results can easily be extended to the case where there are no self-interactions.

Research supported in part by AFOSR MURI project #FA9550–09–1–0538 and the DARPA Physical Intelligence program (contract #HR0011–10–1–0009). M.J. Fox is supported by the Department of Defense through the National Defense Science & Engineering Graduate Fellowship Program.

M.J. Fox and J.S. Shamma are with the School of Electrical and Computer Engineering, College of Engineering, Georgia Institute of Technology
{mfox, shamma}@gatech.edu

This model can be augmented to accommodate differing weights for different symbols and events [8] although we do not consider this here. Characterization of various static equilibria for this model are carried out in [10], and corresponding dynamic models are considered in [9]. A discussion of robustness with respect to the specified learning dynamics can be found in [7]. We have up until now left the computation of $(\bar{P}, \bar{Q})$ from the joint strategy intentionally vague so that the same model can be used in both the infinite and finite population settings. We first describe the infinite population case.

1) Infinite Populations: The standard technique for modeling infinite populations is to consider a continuous mass of players [14]. There are n^2m^2 languages in $\mathcal{L}_{m\times n}$ so we define the population state space as $X = \mathbb{S}^{n^2m^2}$ where $\mathbb{S}^r$ is the r-dimensional simplex. We confer any ordering on $\mathcal{L}_{m\times n}$ so that each element x_i of a state $x \in X$ gives the fraction of the population that speaks a particular language. It follows that the subscripts (P_i, Q_i) refer to the i'th language in $\mathcal{L}_{m\times n}$ (not the i'th player) in this setting and similarly for the utilities u_i. We can then compute

$$(\bar{P}, \bar{Q}) = (\sum_{i=1}^{m^2n^2} x_i P_i, \sum_{i=1}^{m^2n^2} x_i Q_i),$$

the average language in the population at large. The standard evolutionary dynamic for studying games of this type is the replicator dynamic

$$\begin{aligned} \dot{x_i} &= x_i[u_i((P_i, Q_i), (\bar{P}, \bar{Q})) - \bar{u}((\bar{P}, \bar{Q}))] \\ &= x_i[\frac{1}{2}\mathbf{Tr}(P_i\bar{Q}) + \frac{1}{2}\mathbf{Tr}(\bar{P}Q_i) - \mathbf{Tr}(\bar{P}\bar{Q})] \end{aligned}$$

for $i = 1, ..., n^2m^2$. The term $\bar{u}((\bar{P}, \bar{Q})) = \mathbf{Tr}(\bar{P}\bar{Q})$ is the payout to the average of the population when it plays against itself. We will use this quantity as our measure of social welfare. The replicator dynamic is imitative: an unused strategies is never subsequently taken up. It follows that each vertex of the simplex is a rest point of the dynamic. What is surprising about the behavior of this system is that there are many neutrally stable strategies (sometimes referred to as weak evolutionarily stable strategies) where social welfare is not maximized that the system will converge to from a set of initial conditions with non-zero Lebesgue measure [12]. This is troubling for proponents of the simple proto-languages explanation of language origins. The retort is that small populations, where mutations can impact the population state, were integral to the formation of the first proto-languages.

2) Finite Populations: In the finite case, we consider N players and a population state space $X = \mathcal{L}_{m\times n}^N$. For the population state $x \in X$ we let $x_i = (P_i, Q_i)$ refer to the language of player i. We can compute

$$(\bar{P}, \bar{Q}) = (\frac{1}{N}\sum_{i=1}^{N} P_i, \frac{1}{N}\sum_{i=1}^{N} Q_i).$$

We reiterate that in this setting the subscript in x_i refers to the player while in the infinite population setting it refers to the language.

One issue with analyzing the language game in finite populations is that there are many different ways to generalize replicator dynamics and evolutionarily stable strategies (the associated static equilibrium concept) to finite populations (see for instance [15]). One particular approach [6] is to consider the limit of weak selection where the contribution of utility to an otherwise uniform reproductive fitness is taken to zero. For some analytical results associated with this solution concept, see for instance [16]. This is the approach taken in [11]. In that model, one player is selected at random proportional to its fitness and then a second randomly chosen player adopts the first player's language. It is shown that, in the limit of weak selection, population states that maximize social welfare are the only states for which no mutant strategy has a fixation probability higher than $1/N$. This analysis is used to argue that evolution directs the system towards linguistic coherence. However, it is clear that this particular model as specified will not, in general, converge to a socially efficient state with high probability. Such would require analyzing a system that exhibits stronger selection—this is the idea that is pursued in this paper.

Specifically, in Section III we propose a model of reproduction in populations in which a randomly selected individual adopts the language of one of the players that has the current highest utility. That is, unless a mutation occurs with probability ϵ in which case a random language from $\mathcal{L}_{m\times n}$ is adopted. We analyze this model in the small mutation rate limit. The resulting prediction of linguistic coherence is in the form of stochastic stability, a concept introduced to study the evolution of social conventions, but not previously suggested in relation to the language game. We review stochastic stability in Section II. Since stochastic stability is sensitive to the manner in which the mutations are applied [17], we also discuss variations on our model in order to support the view that our results are not especially sensitive to specific features of the model (Section IV). This paper makes three novel contributions: we analyze a stochastic, finite population model of the language game exactly for the case of strong selection, we draw a connection between the study of the evolution of social conventions and language evolution, and we suggest that non-equilibrium models like our own are adequate to explain the observed drift in languages over time.

In the next section, we briefly review the concept of stochastic stability.

II. Stochastic Stability

This introduction to the notion of stochastic stability will draw heavily from the presentation of Young [18] in the context of social conventions. We will develop these concepts here with an eye for brevity. We will consider a Markov process P^0 on a finite state space Z. We will restrict our interest to perturbations to this process of a specific form, defined below.

*Definition 2.1: Let P^ϵ be a Markov process on Z for each $\epsilon \in (0, \bar{\epsilon}]$. The process P^ϵ is a **regular perturbed Markov process** if P^ϵ is irreducible and aperiodic for every $\epsilon \in (0, \bar{\epsilon}]$*

and for each $z, z' \in Z$ we have

$$\lim_{\epsilon \to 0} P^{\epsilon}_{zz'} = P^{0}_{zz'},$$

and if $P^{\epsilon}_{zz'} > 0$ for some $\epsilon > 0$, then

$$0 < \lim_{\epsilon \to 0} P^{\epsilon}_{zz'} / \epsilon^{r(z,z')} < \infty$$

for some $r(z, z') \geq 0$.

The value $r(z, z') \in \mathbb{R}$ is called the *resistance* of the transition $z \to z'$. Clearly, $r(z, z')$ must be uniquely defined in order to satisfy the condition. Also, $P^{0}_{zz'} > 0$ if and only if $r(z, z') = 0$. That is, transitions that occur with non-zero probability under P^0 have zero resistance. Transitions that never occur can be considered as having infinite resistance so that $r(z, z')$ is always defined.

For each ϵ, there is a unique stationary distribution, μ^{ϵ}, associated with P^{ϵ} (by its irreducibility and aperiodicity). We can now formally define stochastic stability.

*Definition 2.2: A state z is **stochastically stable** (Young, 1993) if*

$$\lim_{\epsilon \to 0} \mu^{\epsilon}(z) > 0.$$

It has been shown elsewhere that the above limit exists for every z so that every regular perturbed Markov process has at least one stochastically stable state. These states are the ones that the system spends most time in over the long run when ϵ is small. It should be noted that the stochastically stable states correspond to the perturbed process P^{ϵ}. That is, which states survive in the presence of the perturbations will depend on how the perturbations are introduced. It is possible to arrive at different stochastically stable states for the same process P^0 by applying the perturbations differently. Also, the stochastically stable states correspond to the limiting case of ϵ approaching zero, and are not always particularly likely to be observed when ϵ is not small. Next we will describe how to compute the stochastically stable states.

A. Resistance Trees

A recurrent class of a Markov process is a set of states such that from any state in the set one can reach any other state in the set in finite time with positive probability, and no state outside the set is accessible from any state inside it. Let P^0 have K recurrent classes $E_1, E_2, ..., E_K$. We will define for every distinct pair of recurrent classes E_i and E_j, $i \neq j$, a sequence of states $\zeta = (z_1, z_2, ..., z_q), z_1 \in E_i, z_q \in E_j$ called an ij*-path*. The resistance of the path is the sum of resistances in the sequence, $r(\zeta) = r(z_1, z_2) + r(z_2, z_3) + ... + r(z_{q-1}, z_q)$. We further denote $r_{ij} = \min r(\zeta)$ as the ij-path with least resistance. r_{ij} is always positive because there cannot be a zero resistance path between two distinct recurrent classes.

Now, for each recurrent class E_j, construct a tree rooted at a vertex j corresponding to E_j. That is, a set of $K - 1$ directed edges such that each $E_i, i \neq j$ is represented by a vertex i and there is a unique directed path from any vertex different from j to j. The resistance of such a tree is the sum of the resistances r_{ij} on the $K - 1$ edges. The stochastic potential γ_j of the recurrent class E_j is the minimum resistance among all such trees rooted at j. We expect the recurrent classes of minimum stochastic potential to be the most likely when ϵ is small. This result has been formalized [19] as follows:

Theorem 2.1: Let P^{ϵ} be a regular perturbed Markov process, and let μ^{ϵ} be the unique stationary distribution of P^{ϵ} for each $\epsilon > 0$. Then $\lim_{\epsilon \to 0} \mu^{\epsilon} = \mu^0$ exists, and μ^0 is a stationary distribution of P^0. The stochastically stable states are precisely those states that are contained in the recurrent class(es) of P^0 having minimum stochastic potential.

Next we derive a bound on the stochastic potential of a state based on the construction of greedy, or myopic, forests. The bound will be tight for the models we analyze below.

B. Myopic Forests

In this section we introduce a lower bound on the stochastic potential of a recurrent class based on myopic forests. In the case that a myopic forest can be constructed that is itself a resistance tree, the bound is tight and the potential is the minimum over all resistance trees for that recurrent class. A tree has minimum resistance when the sum of all the resistances is minimum. A myopic forest minimizes the resistance of each outgoing edge individually without any connectedness constraint[1].

Lemma 2.1: Let P^{ϵ} be a regular perturbed Markov process with $E_1, E_2, ..., E_K$ the recurrent classes of P^0, then for any recurrent class j we have

$$\gamma_j \geq \sum_{i \neq j} \min_{k \neq i} (r_{ik}),$$

and the relationship is satisfied with equality whenever there exists a myopic forest $(\{1, ..., K\}, \{(ik) : k \in \textbf{\textit{argmin}}_{k \neq i} r_{ik}\})$ that is a tree rooted at j.

Next we present our dynamic model.

III. The Dynamic Model

The N players play the language game with the following model of reproduction. At each time t select an agent i at random according to some distribution $F(x)$ satisfying $\mathbf{Pr}[F(x) = i] > 0 \; \forall i \in \{1, ..., N\}, x \in X$. Let

$$x_i[t+1] = \begin{cases} x_{\hat{k}}, & w.p. \quad 1 - \epsilon \\ \textbf{rand}(\mathcal{L}), & w.p. \quad \epsilon, \end{cases}$$

where $\hat{k} = \textbf{argmax}_k u_k((P_k[t], Q_k[t]), (\bar{P}[t], \bar{Q}[t]))$, and $\textbf{rand}(\mathcal{L})$ refers to the language given by sampling from the set of possible languages uniformly. Furthermore let

$$x_j[t+1] = x_j(t) \quad \forall \quad j \neq i.$$

In words, we select a random agent and assign him the language of an individual with a utility that is currently highest, or with small probability we assign a random language instead. This dynamic model gives a perturbed Markov processes $\mathcal{P}_{m \times n, N}$ for particular values of m, n, N. We call a state *homogeneous* if for some $l \in \mathcal{L}_{m \times n}$ we

[1]We omit most proofs from this manuscript. An extented online version with full proofs can be accessed from the author's website.

have $x = (l, l, ..., l)$. Clearly the absorbing states of the unperturbed process are precisely the homogeneous states. We next compute the stochastically stable states of this process, first for the case where $m = n$ and then for $m > n$ ($n > m$ is then implied by symmetry). Recall that, in the long run, the process spends an arbitrarily large proportion of its time in the stochastically stable states as ϵ goes to zero.

A. The $m = n$ Case

We will want to make use of the bound we introduced in Lemma 3.1. In order to do so we must compute some minimum resistances from homogeneous states. First consider homogeneous states that maximize linguistic coherence. These are the states that satisfy $\mathbf{Tr}(\bar{P}\bar{Q}) = n$. This condition implies that $\mathbf{Tr}(P_i Q_i) = n \; \forall i \in \{1, ..., N\}$. We call languages satisfying this condition *aligned*. We call the homogeneous states corresponding to aligned languages *optimal*. The next lemma characterizes the minimum resistance from optimal states.

Algorithm 1 `Repair`(P, Q)

1: $\hat{Q}^0 \leftarrow P^{\mathrm{T}}$
2: $\mathcal{K}_2 \leftarrow \{k \in \{1, ..., n\} : \sum_j \hat{Q}^0_{kj} \geq 2\}$
3: $\hat{\mathcal{K}}_2 \leftarrow \{k \in \mathcal{K}_2 : \sum_j \hat{Q}^0_{kj} Q_{kj} = 1\}$
4: **let** $q^k = (q^k_1, ..., q^k_m) = e_x$ s.t. $\hat{Q}^0_{kx} = 1$ for each $k \in \mathcal{K}_2 - \hat{\mathcal{K}}_2$
5: $\hat{Q}^1_{ij} \leftarrow \begin{cases} \hat{Q}^0_{ij}, & i \notin \mathcal{K}_2 \\ Q_{ij}, & i \in \hat{\mathcal{K}}_2 \\ q^i_j, & i \in \mathcal{K}_2 - \hat{\mathcal{K}}_2 \end{cases}$
6: **let** Q' be any binary, row stochastic, and column sub-stochastic $n \times m$ matrix satisfying $Q'_{ij} \geq \hat{Q}^1_{ij} \; \forall \, i, j$
7: **let** P' be any binary, row stochastic $m \times n$ matrix satisfying $P'_{ij} \geq Q'_{ji} \; \forall \, i, j$
8: **return** (P', Q')

Lemma 3.1: When $m = n$ the minimum resistance from an optimal state to any other state is $\lceil N/2 \rceil$ and this is achieved by any other optimal state.

Proof: Suppose a majority of agents speak optimal language l_0 in state x. Furthermore, suppose there are K other languages in x, referred to as $l_k, k \in \{1, 2, ..., K\}$. Note that previously we used l_k to refer to the language of agent k. Let m_k be the number of agents speaking l_k in x. Suppose agent i speaks $l_0 = (P_0, Q_0)$ and agent j speaks $l_r = (P_r, Q_r), r \neq 0$. Let e_P (resp., e_Q) be the number of rows that P_0 and P_r (resp., Q_0 and Q_r) differ in. Now compute $u_i((P_0, Q_0), (\bar{P}, \bar{Q})) - u_j((P_r, Q_r), (\bar{P}, \bar{Q}))$

$$
\begin{aligned}
&= \frac{1}{2}\mathbf{Tr}(P_0 \frac{1}{N}\sum_{k=0}^{K} m_k Q_k) + \frac{1}{2}\mathbf{Tr}(\frac{1}{N}\sum_{k=0}^{K} m_k P_k Q_0) \\
&\quad - \frac{1}{2}\mathbf{Tr}(P_r \frac{1}{N}\sum_{k=0}^{K} m_k Q_k) - \frac{1}{2}\mathbf{Tr}(\frac{1}{N}\sum_{k=0}^{K} m_k P_k Q_r) \\
&= \frac{1}{2}\mathbf{Tr}((P_0 - P_r)\frac{1}{N}\sum_{k=0}^{K} m_k Q_k) \\
&\quad + \frac{1}{2}\mathbf{Tr}(\frac{1}{N}\sum_{k=0}^{K} m_k P_k (Q_0 - Q_r)) \\
&= \frac{1}{2}\mathbf{Tr}((P_0 - P_r)\frac{1}{N} m_0 Q_0) + \frac{1}{2}\mathbf{Tr}(\frac{1}{N} m_0 P_0 (Q_0 - Q_r)) \\
&\quad + \frac{1}{2}\mathbf{Tr}((P_0 - P_r)\frac{1}{N}\sum_{k=1}^{K} m_k Q_k) \\
&\quad + \frac{1}{2}\mathbf{Tr}(\frac{1}{N}\sum_{k=1}^{K} m_k P_k (Q_0 - Q_r)) \\
&\geq \frac{1}{2}\frac{1}{N} e_P m_0 + \frac{1}{2}\frac{1}{N} e_Q m_0 - \frac{1}{2}\frac{1}{N} e_P \sum_{k=1}^{K} m_k \\
&\quad - \frac{1}{2}\frac{1}{N} e_Q \sum_{k=1}^{K} m_k \\
&= \frac{1}{2}\frac{1}{N}(e_P + e_Q)(m_0 - \sum_{k=1}^{K} m_k) > 0
\end{aligned}
$$

The second-to-last inequality follows from the fact that $|\mathbf{Tr}((A - B)C)|$ is always less than the number of mismatched rows among A, B when $A, B, C \in \mathbb{B}^{m \times n}$. The last inequality follows from the assumption that a majority of individuals speak l_0. Since l_0 speakers have the highest utility as long as they have a majority, we require enough mutations to institute a new majority aligned language, which is $\lceil N/2 \rceil$ starting from the state homogeneous in l_0. Afterwards, the state can become homogeneous in the new language without resistance. ■

From states that are not optimal we can always reach some optimal state with resistance equal to one. The optimal state we can reach depends on the state that we start from. This is because any language that is not aligned has a corresponding aligned language that, when introduced via mutation, has utility as least as high as the incumbent. We use the `Repair` algorithm to find this language. We will use this algorithm for the $m \neq n$ case as well, so we assume w.l.o.g. below that $m \geq n$. The idea of `Repair` is to construct a new Q-matrix that is close to P^{T}. This is because $\mathbf{Tr}(PP^{\mathrm{T}}) = m$, so P^{T} is always "close" to a good hearer matrix with respect to P-speakers. We say "close" because P^{T} will not, in general, be row stochastic. We begin by massaging $\hat{Q}^0 = P^{\mathrm{T}}$ into a row sub-stochastic matrix $\hat{Q}^1$ (lines 1-5). The rows of $\hat{Q}^0$ that sum to 1 are left unchanged. We distinguish two cases for when a row of $\hat{Q}^0$ sums to more than 1. That is, rows

that give a positive dot product with their corresponding row in Q (lines 2-3) and those that do not. This is because we will eventually obtain our speaker matrix P' from our hearer matrix Q' so we must anticipate how well this matrix speaks to Q. Any of the 1's in a row of $\hat{Q}^0$ suffice to correctly associate the corresponding symbol received via P, however, at most one of these 1's is in the same position as the 1 in the corresponding row of Q. Therefore, when it is possible to satisfy both interests we do so. The matrix $\hat{Q}^1$ is then row sub-stochastic— it may still contain zero rows. In line 6 we add 1's to these zero rows to get a proper hearer matrix Q' that is column sub-stochastic. Since Q' is column sub-stochastic its transpose is row sub-stochastic so line 7 merely adds 1's to zero rows in order to get a row stochastic P'.

Example: `Repair`

$$(P,Q) = \left(\begin{bmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{bmatrix} \right),$$

$$\Rightarrow \texttt{Repair}(P,Q) = \left(\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \right). \S$$

The next two lemmas establish the behavior of `Repair`.

Lemma 3.2: (P',Q') =`Repair`(P,Q) is an aligned language for any (P,Q).

Lemma 3.3: Every suboptimal homogeneous state can reach an optimal state with resistance one.

We can now give the main result for the $m = n$ case.

Theorem 3.1: For any $N \geq 3$ and any $m \geq 2$ the stochastically stable states of the process $\mathcal{P}_{m\times m,N}$ are the optimal states.

Next, we develop a similar result for the case of $m > n$.

B. The $m > n$ Case

We will again utilize Lemma 3.1. The minimum resistance targets from sub-optimal states are the same as in the $m = n$ case. However, the minimum resistance targets from optimal states are different in this case.

Lemma 3.4: The minimum resistance between an optimal state and any other state is 1, and this is achieved by an optimal language.

Every homogeneous (absorbing) optimal state can transition to *some* other homogeneous optimal states with resistance 1. Are there any optimal states that cannot be reached via a sequence of transitions through optimal states, each having resistance 1? We answer this question in the affirmative with a constructive algorithm called `Path`, whose details are ommited from this manuscript.

Lemma 3.5: The `Path` algorithm takes two aligned languages, one initial and one final, and returns a sequence of aligned languages linking the associated initial and final optimal states via transitions of resistance 1.

We can now state our main result for the $m > n$ case.

Theorem 3.2: For any $N \geq 3$ and any $m > n \geq 2$ the stochastically stable states of the process $\mathcal{P}_{m\times n,N}$ are the optimal states.

In the next section, we show that the results of Theorems 3.1 and 3.2 also apply to several different variations on the reproduction dynamics.

IV. Other Dynamic Models

The reproduction dynamics of $\mathcal{P}_{m\times n,N}$ rely on an important assumption: only the individuals with the highest utility are able to spread their language to others (other than via mutation). Thus, the dynamic can be considered strongly selective. In this section we introduce three variations on the reproduction dynamics that relax this assumption. These are twice-perturbed dynamics, pairwise competition dynamics, and pairwise competition on a fixed competition graph. In each case we argue informally that the dynamics give the same set of stochastically stable states.

A. Twice-perturbed Dynamics

In a more realistic model we should expect to see languages not giving the highest utility reproducing themselves some of the time. We can model this phenomenon as a second perturbation to the process. At time t, select a language

$$(P',Q') = \begin{cases} x_{\hat{k}}[t], & w.p. \quad 1-\rho\epsilon \\ x_{\mathbf{rand}(\{1,\ldots,N\})}[t], & w.p. \quad \rho\epsilon \end{cases}$$

where $\hat{k} = \mathbf{argmax}_k u_k((P_k[t],Q_k[t]),(\bar{P}[t],\bar{Q}[t]))$. Then select an agent i according to F as before. Let

$$x_i[t+1] = \begin{cases} (P',Q'), & w.p. \quad 1-\epsilon \\ \mathbf{rand}(\mathcal{L}), & w.p. \quad \epsilon, \end{cases}$$

and let

$$x_j[t+1] = x_j[t] \quad \forall \quad j \neq i,$$

where $\rho \in (0, 1/\epsilon)$. In our original dynamics, the perturbations corresponded to mutations. We now have an additional perturbation that corresponds to less fit languages reproducing. This new perturbation can occur with any probability proportional to ϵ. It is easy to verify that these two processes have the exact same set of stochastically stable states. This is because mutation alone still gives a probability of less fit languages reproducing that is proportional to ϵ. Next, we consider a dynamic that relaxes the strength of selection in a different manner.

B. Pairwise Competition Dynamics

Another way to arrive at a weaker form of selection is to suppose that rivalries determine which individuals get the opportunity to reproduce themselves. Thus an individual whose utility is not greater than all others may still spread his language by overtaking a rival with an even smaller utility. At time t, select two players i and j with $i \neq j$ according to a random process over the pairs $F(x[t])$ that is bounded away from zero. Assume without loss of generality that $u_i((P_i[t],Q_i[t]),(\bar{P}[t],\bar{Q}[t])) \geq u_j((P_j[t],Q_j[t]),(\bar{P}[t],\bar{Q}[t]))$. Let

$$x_j[t+1] = \begin{cases} x_i[t], & w.p. \quad 1-\epsilon \\ \mathbf{rand}(\mathcal{L}), & w.p. \quad \epsilon, \end{cases}$$

and let

$$x_k[t+1] = x_k(t) \quad \forall \quad k \neq j.$$

We still compute utilities in the same manner as before. These dynamics can be interpreted as having agents compete over reproductive resources (food, mates, nesting sites, etc.) locally, but with their fitnesses determined from global considerations. Clearly, an agent not speaking the most fit language will often overtake an agent speaking an even less fit agent in this model. It is straightforward to demonstrate that these dynamics give the same resistances as our original reproduction model. Hence they have the same stochastically stable states.

An important feature of pairwise competition is that in the long-run, every player competes with every other player infinitely often. In the next section we consider the opposite-extreme case where the competition graph is fixed for all t.

C. Pairwise Competition Dynamics on a Fixed Graph

Suppose there is a connected graph $G = (\{1, ..., N\}, E)$ describing rivalries amongst the players. Now consider a modified version of pairwise competition dynamics where at each time t we select an edge from E, giving us our two players i and j. Assume that the edges are selected according to any process that gives a probability of selecting a particular edge that is bounded away from zero. Here too, the resistances, and hence the stochastically stable states will be the same as in our original model. The key to seeing this is to note that whenever the state is not homogeneous there is at least one edge at which the players speak different languages. Since pairwise competition only changes the state when two such agents compete, the dynamics have the same qualitative behavior in the long run.

V. DISCUSSION

We analyzed this process separately for the cases where the number of objects and symbols agree and disagree. In the more natural setting where the number of objects and symbols disagree we showed that we could transit between any two optimal states through a sequence of optimal states requiring only one mutation per transition. This (along with the non-equilibrium nature of the process) concords with the observed phenomenon of drift in languages. That is, languages seem to change over time (see for instance, [20]) in a manner that is neutral with respect to the expressiveness of the language. The presence of synonyms and homonyms, exploited in our `Path` algorithm, seems a reasonable mechanism for this action.

A. Future Directions

We considered only the situation where mutations select from the entire set of possible languages. It is more natural to consider point mutations. That is, when a mutation occurs the mutating individual modifies only a single row of one of the matrices he would have adopted without mutation. This case requires some additional arguments that we present in an upcoming paper.

A second feature of the model that can be generalized is the form of the utility. Although we considered more local interaction frameworks in the form of pairwise competition and pairwise competition on a graph, we still computed the utility in a manner reflecting a global interaction. It is possible to instead compute each agents utility based on their ability to communicate with some subset of the total population. This subset could come from either some fixed, exogenous graph or some endogenous considerations. An interesting question that emerges when considering these circumstances is the problem of linguistic diversity. What conditions are needed for heterogeneous states to persist in the population with non-vanishing frequency? Can we quantify network effects on social welfare? These are among the interesting questions that can be studied by considering generalizations to the utility functions of this game that move away from the "everyone talks to everyone" paradigm. These extensions are being pursued by the authors.

REFERENCES

[1] M. Balter, "Animal Communication Helps Reveal Roots of Language," *Science*, vol. 328, no. 5981, pp. 969–971, 2010.

[2] L. Cavalli-Sforza, "Genes, peoples and languages," *Proc. Natl Acad. Sci. USA*, vol. 94, pp. 7719–7724, 1997.

[3] M. Nowak, N. Komarova, and P. Niyogi, "Computational and evolutionary aspects of language," *Nature*, vol. 417, pp. 611–17, 2002.

[4] S. G. Ficici and J. B. Pollack, "Coevolving communicative behavior in a linear pursuer-evader game," in *Proceedings of the Fifth International Conference of the Society for Adaptive Behavior*, K. Pfeifer, Blumberg, Ed. Cambridge: MIT Press, 1998.

[5] C. H. Yong and R. Miikkulainen, "Coevolution of role-based cooperation in multi-agent systems," Department of Computer Sciences, The University of Texas at Austin, Tech. Rep. AI07-338, 2007.

[6] M. Nowak, A. Sasaki, C. Taylor, and D. Fudenberg, "Emergence of cooperation and evolutionary stability in finite populations," *Letters to Nature*, pp. 646–650, April 2004.

[7] S. Huttegger, "Robustness in signaling games," *Philosophy of Science*, vol. 74, pp. 839–847, 2007.

[8] G. Jager, "Evolutionary stability conditions for signaling games with costly signals," *Journal of Theoretical Biology*, vol. 253, no. 1, pp. 131 – 141, 2008.

[9] M. Nowak, J. Plotkin, and D. Krakauer, "The evolutionary language game," *Journal of Theoretical Biology*, vol. 200, no. 2, pp. 147 – 162, 1999.

[10] P. Trapa and M. Nowak, "Nash equilibria for an evolutionary language game," *Journal of Mathematical Biology*, vol. 41, pp. 172–188, 2000, 10.1007/s002850070004.

[11] C. Pawlowitsch, "Finite populations choose an optimal language," *Journal of Theoretical Biology*, vol. 249, no. 3, pp. 606 – 616, 2007.

[12] ——, "Why evolution does not always lead to an optimal signaling system," *Games and Economic Behavior*, vol. 63, no. 1, pp. 203 – 226, 2008.

[13] D. Lewis, *Convention: A Philosophical Study*. Harvard Univ. Press, Cambridge, MA., 1969.

[14] W. H. Sandholm, *Population Games and Evolutionary Dynamics*. MIT Press, 2010.

[15] J. Maynard Smith, "Can a mixed strategy be stable in a finite population?" *J. Theor. Biol.*, vol. 130, pp. 247–251, 1988.

[16] T. Antal, A. Traulsen, H. Ohtsuki, C. E. Tarnita, and M. A. Nowak, "Mutation-selection equilibrium in games with multiple strategies," *Journal of Theoretical Biology*, vol. 258, no. 4, pp. 614 – 622, 2009.

[17] J. Bergin and B. Lipman, "Evolution with state-dependent mutations," *Econometrica*, vol. 64, no. 4, pp. 943–56, July 1996.

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

[19] ——, "The evolution of conventions," *Econometrica*, vol. 61, no. 1, pp. 57–84, January 1993.

[20] O. Jespersen, *Progress in Language with Special Reference to English*. London: Swan Sonnenschein & Co., 1894.