Journal of Theoretical Biology 264 (2010) 104–118

Contents lists available at ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

Eavesdropping and language dynamics

Natalia L. Komarova [a,*], Simon A. Levin [b]

[a] Department of Mathematics, University of California Irvine, 540J Rowland Hall, Irvine, CA 92697, USA
[b] Department of Ecology and Evolutionary Biology, Princeton University, 106 Guyot Hall, Princeton, NJ 08544, USA

ARTICLE INFO

Article history:
Received 22 September 2009
Received in revised form
25 November 2009
Accepted 11 January 2010
Available online 20 January 2010

Keywords:
Language evolution
Animal communication
Synonyms
Communication games
Agents
Language dynamics equation
Alarm calls

ABSTRACT

Communication in nature is not restricted to the transmitter–receiver pair. Unintended listeners, or eavesdroppers, can intercept the signal and possibly utilize the received information to their benefit, which may confer a certain cost to the communicating pair. In this paper we explore (computationally and mathematically) such situations with the goal of uncovering their effect on language evolution. We find that in the presence of eavesdropping, languages exhibit a tendency to become more complex. On the other hand, if eavesdroppers belong to a different (competing) population, the languages used by the two populations tend to converge, if the cost of eavesdropping is sufficiently high; otherwise the languages synchronize. These findings are discussed in the context of animal communication and human language. In particular, the emergence of synonyms is predicted. We demonstrate that a small associated cost can suppress synonyms in the absence of eavesdropping, but that their likelihood increases strongly with the probability of eavesdropping.

1. Introduction

Communication is widespread in nature. Mating calls, alarm calls warning others about predators, communications about the location and quality of food sources—many species use highly specialized vocalizations or other specific means to convey different types of information. Different species develop their own unique ways of communication. It is not surprising that species that share habitats learn to recognize and utilize each other's calls.

Eavesdropping in ecological systems has attracted attention of many researchers (see e.g. Magrath et al., 2009; Searcy and Nowicki, 2005, and the references therein). Vervet monkeys respond to the alarm calls of superb starlings (Hauser, 1988); mongooses eavesdrop on hornbill calls (Rasa, 1983). The chickadee acts as a guard for anywhere between 24 and 50 other bird species, who overhear their alarm calls and gather to mob the predator away from the site (Templeton and Greene, 2007). Galapagos marine iguanas, though mute, recognize and utilize the alarm calls of the Galapagos mocking birds (Vitousek et al., 2007).

The above examples of eavesdropping show that coexisting species learn to communicate by recognizing each other's signals. These species may compete with each other for space and resources (Schmidt and Ostfeld, 2008), or co-operate (Templeton and Greene, 2007). Other examples of inter-species communication come from interactions between prey and predators. There are instances where prey learns to utilize the language of the predator, and where predator communicates with prey. An example of the former type of behavior is given by ground squirrels who learned to use an infrared signal to deter their predator, rattlesnakes (Rundus et al., 2007). The latter type of interactions is exemplified by photuris fireflies, who mimic female photinus fireflies by scent and glow patterns in order to lure photinus males, which they then devour (Lloyd, 1975).

The term "eavesdropping" (McGregor, 1993; McGregor and Peake, 2000) used to describe these and other interactions in biology, clearly comes from human experience, where humans try to overhear each other's talking, and find out information they can use to their benefit. Eavesdropping can be essential when two groups of people are at war, or in competition with each other. More generally, it is a strategy that can be utilized by members of one group when the other group is trying to hide something. There are many situations where eavesdropping (and consequently, strategies to avoid eavesdropping) can be important, ranging from tribal conflicts to the usage of slang by criminals or youth, who want to hide certain information from the rest of the population.

While a growing body of literature exists which shows various examples of eavesdropping, there have been no mathematical studies of its implications for the development of communication systems. Most work on the evolution of signalling systems focuses on within-species interactions. An important concept of

* Corresponding author.
E-mail address: komarova@math.uci.edu (N.L. Komarova).

0022-5193/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jtbi.2010.01.010

communication networks was introduced by McGregor and Dabelsteen (1996) and McGregor and Peake (2000), but no quantitative analysis of its consequence for language evolution has been performed. How does the presence of eavesdroppers affect the direction of language evolution, both within one species, and interspecifically? Can one species adapt its language to inhibit eavesdropping? Does this have any consequences for language complexification, or diversification of languages?

In the present paper we address these questions in the context of human and animal communication. We consider an agent-based model of individuals who play communication games and whose fitness depends on their communication. Individuals can eavesdrop on each other which results in the fitness changes for some or all parties involved. The resulting language dynamics is studied numerically. The key results are also studied analytically by using simple ordinary differential equation models based on language dynamics equations.

The rest of the paper is organized as follows. In Section 2 we present our modeling approach. In Section 3 we discuss language complexification. We discover two mechanisms of complexification: (i) due to an asymmetrical payoff for speaking and for receiving, and (ii) due to eavesdropping. In Section 4 we talk about language diversification in the context of eavesdropping. In Section 5 we look at languages as signal-meaning mappings, and study both complexification and diversification of languages; in particular, we discuss the emergence of synonyms as a result of eavesdropping. Section 6 is reserved for discussion and conclusions.

2. The modeling approach

2.1. Populations and languages

Consider a population of n agents. Each agent is equipped with a "language". As the initial condition, we can assume that the languages are assigned at random, or that everybody speaks the same language. Each agent is assigned some communication-independent fitness value, F_0, which is assumed to be the same for all.

There are m possible languages, L_i with $1 \le i \le m$. A similarity matrix, $\{s_{ij}\}$, specifies the probability that a carrier of language L_j will understand an utterance spoken in language L_i. In general, $s_{ij} \ne s_{ji}$, and $0 \le s_{ij} \le 1$. In Nowak et al. (2001) and Komarova et al. (2001), a fully symmetrical similarity matrix was used, such that $s_{ij} = a < 1$ for all $i \ne j$, and s_{ii}=1 for all i. Another model was to take the off-diagonal coefficients s_{ij} to be random numbers uniformly distributed between 0 and 1.

Here we will be interested in models with a certain structure imposed on the set of languages. This structure manifests itself in the form of the matrix $\{s_{ij}\}$. Several specific models of language similarity are presented below.

Model I: *Inclusive sets*. Let all the languages be numbered from 1 to m, such that language L_{i+1} includes language L_i. This way, we have

$$s_{i,j} = \begin{cases} 1, & j \ge i, \\ s_{ij} < 1, & j < i, \end{cases} \quad s_{ij} < s_{ik} \text{ if } j < k < i,$$

which imposes a simple hierarchical structure on the set of languages. In this (very restrictive) sense, language L_{i+1} is more complex than language L_i. As an example, we will be using the matrix $\{s_{ij}\}$:

$$s_{ij} = \begin{cases} 1, & j \ge i, \\ e^{-\gamma|i-j|}, & j < i, \end{cases} \tag{1}$$

where γ measures by how much each consecutive language is more complex than the previous one. In this model, s_{ii}=1 for all i.

Model II: *Binary strings*. Another model of language similarity is inspired by thinking of languages as sets of rules. Let us suppose that while some rules are fixed and common to all languages, there are k other rules that can vary from language to language (we are assuming that these rules are independent of each other). For simplicity, suppose that each rule can exist in two versions, 0 and 1. Then all languages can be represented as binary strings of length k. For example, with k = 2, we have four possible languages: (00), (01), (10), and (11).

We assume that version 1 of a rule includes version 0. That is, someone with version 1 can understand the version 0 of the rule, but not the other way around. In the example above, language (11) understands all the other languages, languages (10) and (01) understand 50% of each other's utterances, and all the languages understand language (00) completely.

We will call the "rank" of a language the total number of "ones" in the binary representation of this language. For example, language (00) has rank 0 and language (11) has rank 2, the maximum possible rank with k = 2. A language of a higher rank is considered to be more complex than a language of a lower rank.

Once this combinatorial structure is assumed, the coefficients of the similarity matrix can be calculated as follows:

$$s_{ij} = 1 - d_{ij}/k,$$

where d_{ij} is the number of positions where language j has a zero and language i has a one. Again, in this model, s_{ii}=1 for all i.

Model III: *A non-hierarchical chain*. In this model, we envisage an ordered chain of m languages such that any two consecutive languages are close together, and the difference between languages i and j grows with $|i-j|$. This corresponds to a similarity matrix which has ones on the diagonal, and decreasing elements off the diagonal.

2.2. The communication game

At each round of the communication game, n_{update} pairs of agents are picked at random from the population. For example, if n_{update}=n, then as many pairs are picked as there are people in the population (this does not mean that everybody gets picked, and some agents are picked more than once). Each pair exchanges information, such that one of the agents is a speaker and the other is a listener.

After each information exchange event, the fitness values of the participating individuals are updated. The language-dependent fitness correction is defined in the following way. Suppose, the two agents that communicate are agents i and j, where i speaks and j listens. They use languages $L(i)$ and $L(j)$, respectively.[1] Then we have the fitness increment for the speaker,

$$\Delta f_i = G_{send} s_{L(i),L(j)},$$

and for the listener,

$$\Delta f_j = G_{rec} s_{L(i),L(j)}.$$

Here, G_{send} and G_{rec} are constants defining the payoff to transmit and to receive information, respectively. In the most general case we assume that the interests of the speaker and the receiver do not necessarily coincide (Seyfarth and Cheney, 2003), so the two constants G_{send} and G_{rec} may be different.

[1] Here, $L(i)$ stands for "the language of agent i" with $i \in [1,n]$, rather than "language number i" with $i \in [1,m]$; the latter is denoted by L_i.