ARTICLE IN PRESS

Available online at www.sciencedirect.com

SCIENCE DIRECT

ELSEVIER

Journal of Theoretical Biology ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Journal of Theoretical Biology

www.elsevier.com/locate/yjtbi

Letter to the editor

Network topology and self-consistency in language games

The evolutionary dynamics of language has attracted considerable attention over the last years. Different approaches have been taken, from embodied communicating robotic agents (Steels, 2003, 2000; Hashimoto, 1997) to abstract models of signal-object associations (Komarova and Niyogi, 2004). In the later case, strong assumptions are typically made in order to reduce the potential complexity of the game dynamics, in such a way that analytic results can be extracted. A relevant example in this context is a family of language games involving a well-defined pay-off defined in terms of the structure of the lexical matrix where all agents interact with each other (Nowak and Krakauer, 1999, Nowak et al., 1999; Komarova and Niyogi, 2004; Komarova and Nowak, 2001). This type of games leads eventually to a shared code among agents.

All these models consider a scenario of agent interactions where all-to-all exchanges occur. More precisely, the fitness function measuring the pay-off associated to proper communication is computed by matching the performance of each player with all the others. Real networks involving social interactions do not need to follow such a rule. Actually, it seems clear from available data that social networks are typically sparse and have small world structure (Watts and Strogatz, 1998). Small path lengths between agents and large clustering are the two essential characteristics of these graphs. Here clustering refers to the presence of order at the local network level, defined in terms of the probability of finding triangles (Dorogovtsev and Mendes, 2002). Network topology largely influences the way information or epidemics spreads through the community, and both local and global properties can constrain the way communication develops and evolves (Solé et al., 2002). It might actually pervade the emergence of complex language traits, such as syntax (Ferrer-Cancho et al., 2005; Solé, 2005).

Here we consider the impact of local network properties in the graph G of agent–agent interactions on language games. As will be shown below, an important assumption of these games, self-consistency, is automatically achieved provided that non-zero clustering is present in a connected network.

Specifically, let us consider a model described by a population of communicating agents, each carrying association matrices linking a set of n signals to a set of m referents (Fig. 1). For our set of agents $\{S_i\}$ $(i = 1, \ldots, N)$ (natural or artificial) we have, for each agent S_i, two matrices $\mathbf{P}^i = (P^i_{kl})$ and $\mathbf{Q}^i = (Q^i_{kl})$. The matrix elements P^i_{kl}, Q^i_{kl} indicate the probabilities that this agent associates object k given the signal l, and to associate the signal l given an object k (here $k = 1, \ldots, n$ and $l = 1, \ldots, m$). The normalizations

$$\sum_{k \leqslant n} P^i_{kl} = 1, \tag{1}$$

$$\sum_{k \leqslant m} Q^i_{lk} = 1, \tag{2}$$

are to be expected. Each agent is thus defined as pair $L_i \equiv L_i(\mathbf{P}^i, \mathbf{Q}^i)$. Agents themselves belong to a network G defined by the set of communicating agents (the nodes) being two agents linked if they exchange information (Fig. 1). In previous studies, global communication is assumed: all agents interact between them and thus the graph describing their communicative exchanges is a complete (fully connected) graph or *clique* (Fig. 2(a)).

A fitness function is defined in order to weight the success of the communication among agents. The total pay-off $F(L_i, L_j)$ associated to a given pair of communicating agents $\{S_i, S_j\}$ is defined as follows:

$$F(L_i, L_j) = \frac{1}{2} \sum_{k \leqslant n} \sum_{l \leqslant m} (P^i_{kl} Q^j_{lk} + P^j_{kl} Q^i_{lk}), \tag{3}$$

which is a symmetric function (i.e. $F(L_i, L_j) = F(L_j, L_i)$) and is such that $F_{max}(L_i, L_j) = \min\{m, n\}$.

A key assumption in previous models of emergence of communication is *self-consistency*: agents understand themselves. This is a sensible assumption, but we can show that there exist configurations able to display the maximum pay-off with non-self-consistent languages: to see this, let us first start with a system composed by two agents. An optimum is obtained provided that either

$$L_i \neq L_j \wedge (\mathbf{P}^j)^{\mathrm{T}} = \mathbf{Q}^i \wedge (\mathbf{P}^i)^{\mathrm{T}} = \mathbf{Q}^j, \tag{4}$$

or alternatively

$$L_i = L_j = L_A \wedge (\mathbf{P}^A)^{\mathrm{T}} = \mathbf{Q}^A \tag{5}$$

and, for any optimum,

$$(\forall l)(\exists k)|(P^i_{kl} = 1 \wedge Q^i_{lk} = 1) \tag{6}$$

(the other matrix elements will be zero, consistently with the normalization condition). The paradox comes here: for the first situation there is no relation between $(\mathbf{P}^i)$ and $\mathbf{Q}^i$

0022-5193/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jtbi.2005.11.025

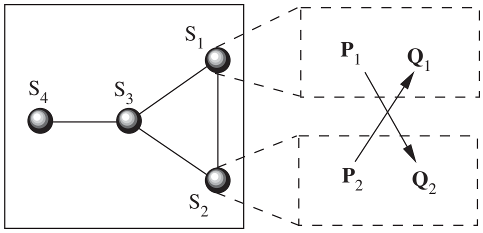

Fig. 1. Agent interaction networks. An example of the agent-agent interaction network defining the topological pattern of possible communication exchanges (left). Each agent has an internal state defined by its **P** and **Q** matrices. Communication interactions (right) are defined as interactions between matrices (see text).

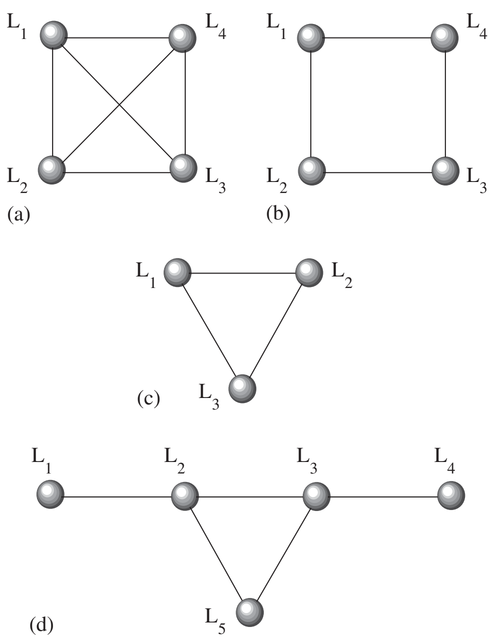

Fig. 2. Some examples of agent graphs G describing the topology of interactions among communicating agents. Here each ball represents a given agent (with the underlying $\mathbf{P}^i$ and $\mathbf{Q}^i$ matrices defining its language) and links connecting two nodes indicate that the game takes place between those two agents: (a) a clique (full-connected) graph with four agents; (b) a square: the smallest loop of a pair number of agents; (c) a triangle: the smallest clique and the smallest loop of an odd number of agents; and (d) a chain of agents with a loop of an odd number of agents.

and thus

$$F(L_i, L_j) = F_{max} \nRightarrow F(L_i, L_i) = F_{max}. \tag{7}$$

This implies that S_i and S_j can understand each other but might not understand themselves: there is no self-consistency.

The so-called lexical matrix has been proposed, among other reasons, as an assumption to avoid such undesirable result (Komarova and Niyogi, 2004). Under such framework, every agent has a lexical matrix $\mu_k = \mu_k(s_i, m_j)$, where $\{s_1 \dots s_n\}$ is the set of all possible signals and $\{m_1 \dots m_m\}$ the set of all possible meanings (Komarova and Niyogi, 2004; Komarova and Nowak, 2001). From the lexical matrix we can derive the P_{ij} and Q_{ji} matrices as follows:

$$P^k_{ij} \equiv \mu_k(s_i|m_j),$$
$$Q^k_{ji} \equiv \mu_k(m_j|s_i). \tag{8}$$

Under these assumptions, self-consistency is guaranteed if the agents reach the maximum pay-off.

The topology of communication exchanges has not being taken into account in previous studies. There is a large number of possible interaction patterns described by the graph G. Some examples (for $N = 4, 5$) are shown in Fig. 2. Here we show that such an assumption might not be required in order to obtain self-consistency, provided that the structure of the agent–agent communication graph has non-zero clustering and is not a bipartite graph (see below). Under these conditions, the maximum pay-off is reached only by self-consistent languages.

Beyond the two-agent system, we can consider a more general situation described by a linear chain of agents using two given languages L_A and L_B. In this case, two neighbouring agents S_i, S_{i+1} on the chain will have a maximum pay-off $F(L_i, L_{i+1}) = F_{max}$ if

$$(\mathbf{P}^{i+1})^{\mathrm{T}} = \mathbf{Q}^i \wedge (\mathbf{P}^i)^{\mathrm{T}} = \mathbf{Q}^{i+1}. \tag{9}$$

This system leads to two possible sets of solutions: either a chain of alternating agents with two languages, i.e.

$$L_A - L_B - L_A - L_B - L_A \cdots$$

as shown in Fig. 3(a), that are not self-consistent, with

$$(\mathbf{P}^B)^{\mathrm{T}} = \mathbf{Q}^A \wedge (\mathbf{P}^A)^{\mathrm{T}} = \mathbf{Q}^B \tag{10}$$

or a completely homogeneous string sharing the same self-consistent language L_A (Fig. 3(b)) with $(\mathbf{P}^A)^{\mathrm{T}} = \mathbf{Q}^A$. If both solutions hold Eq. (6), they are equally good, with $F = \min\{m, n\}$.

The previous example reveals that the problem of self-consistency (present in the first solution) is expected to occur. Let us consider the simplest example: a triangle formed by three interacting agents (in other words, the smallest clique that can be defined, see Fig. 3(c). For this system, it can be easily shown that the only possible solution is

$$L_1 = L_2 = L_3 = L \tag{11}$$

and thus L has to be self-consistent. Now, consider a string of agents following Eq. (9). If L_A and L_B satisfy both Eqs. (4) and (6), the pay-off of all agents is maximal. Here languages are not self-consistent but every agent and the whole system have reached the maximum pay-off (Fig. 3(a)). If we add an agent linking it to a single randomly chosen agent in the chain, the situation does not change. But if we add a new agent interacting with two neighbouring agents on the string (forming a triangle, see Fig. 3(c)) there will be an evaluation of the pay-off like

$$F(L_i, L_j) = F(L_A, L_A) \tag{12}$$

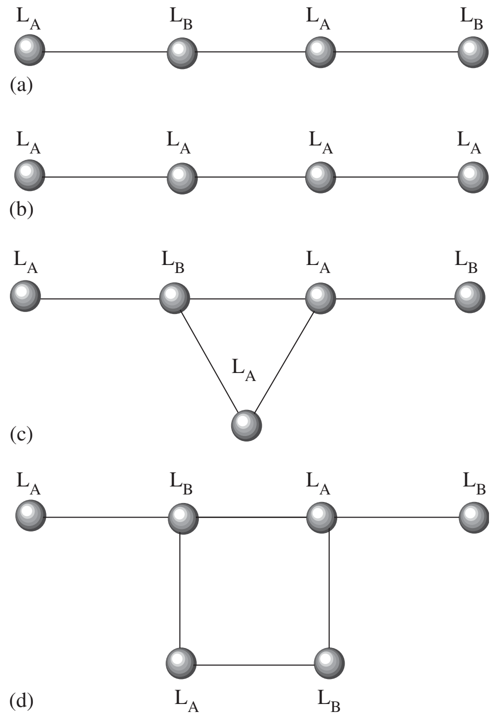

Fig. 3. Different distributions of agent's language in an interaction graph: (a) Non-homogeneous string with two non-self-consistent languages is able to achieve the maximum pay-off; (b) An homogeneous string with one self-consistent language. In terms of pay-off evaluation, this configuration is identical to (a). (c) A chain with an odd loop. The homogeneous distribution of a self-consistent language will be the only set of solutions able to reach the maximum pay-off. (d) If we add a loop of a pair number of agents the two sets of solutions are able to reach the maximum pay-off.

and maximal pay-off will be reached only if L_A is self-consistent. i.e.:

$$F(L_A, L_A) = F_{max} \Longleftrightarrow (\mathbf{P}^A)^{\mathrm{T}} = \mathbf{Q}^A. \tag{13}$$

Note that, if we need L_A to be self-consistent in order to reach the maximum pay-off, the only solution for L_B is $L_B = L_A$. Thus, by adding a triangle to a linear chain of communicating agents, we restrict the set of code configurations that can achieve the maximum and we force the system to exhibit homogeneity and self-consistency.

A generalization of this result can be introduced by using graph colouring (Wilson, 1996, Chapter 6). In general, a given graph is *k-colourable* if we can assign one of k colours to each node such that adjacent nodes have different colours. If G is k-colourable but not $(k-1)$-colourable, we say that G has a chromatic number $\chi(G) = k$. For $k = 2$, a graph G such that can be split into two disjoint sets A and B so that each edge of G links an element of A and another of B is named *bipartite*. If loops of odd number are present in G, the graph cannot be bipartite. Using the property that any bipartite network $\mathcal{G}$ has colouring number $\chi(\mathcal{G}) = 2$ (Bollobás, 1998, p. 146) we can think L_A and L_B as the *colours* used to *paint* the nodes.

Let us assume that the network is not bipartite (i.e. it is connex and it has odd loops), and thus it is such that $\chi(G) > 2$. In such case, two colours are not enough to colour the whole network. This implies that there would be somewhere a pair of connected nodes with the same colour, i.e. a pair L_i–L_i. In a general framework, we can state that, if L_A and L_B are not self-consistent and the network displays $\chi(\mathcal{G}) > 2$, there will not exist any configuration of L_A and L_B able to reach F_{max}. Note that, Fig. 3(a, b) and (d) are 2-*colourable* networks, whereas Fig. 3(c) is a 3-*colourable* network. As a natural consequence, we find that if the clustering coefficient is higher than zero, then only configurations with uniform distribution of a self-consistent language will be able to reach F_{max}, provided that for such a network, $\chi(\mathcal{G}) \geqslant 3$.

Following this reasoning, we can see that the topology of interactions plays a crucial role in order to reach the maximum pay-off with self-consistency and homogeneity. Specifically, it is easy to prove that, if our G graph is tree-like or has no loops with an odd number of nodes (Fig. 3(d)), then the maximum pay-off does not imply self-consistency. On the other hand, if we have at least one triangle (Fig. 3(c)) or one loop with an odd number of agents, then optimality leads to self-consistency and homogeneity.

Our results have implications for both artificial and natural systems. For artificial systems (see for example Steels, 2003) this is a relevant feature: embodied communicating agents should include self-consistency within their embodiment. But self-consistency can be achieved with simple assumptions about connectivity, simplifying the endowment and learning algorithms of agents. For natural systems (including human language) the interpretation of this result has to be done carefully. Here the definition of the lexical matrix μ is justified also as a part of the (underlying) cognitive structure of agents (Komarova and Nowak, 2001; Komarova and Niyogi, 2004). However, small-worldness (and thus non-zero, high clustering) is already present in social networks of both humans and animals (Lusseau, 2003; Newman, 2003) and a priori, specific properties assigned to the lexical matrix are not so relevant. This matrix might be better conceived as a byproduct of evolution, somewhat imposed by the architecture of social networks and not as a primary tool in the endowment of cognitive apparatus of such agents. In mapping this type of games to natural communication phenomena, what is in fact necessary to assume is that there are some computational abilities not changing in time and allowing agents to converge into a common code of signal–meaning association. The evolution and nature of this computational endowment is an open question in both linguistics and biology (Hauser et al., 2002).

A final note concerns the possible set of solutions able to reach the optimum. First, the derivations have been done assuming ideal communication, without environmental noise. It is the simplest assumption to begin with. More realistic approximations should consider noise, signal hierarchies or effort consensus between agents (Ferrer-Cancho and Solé, 2003). A second consideration has to do with the generality of our result, since we have shown that the architecture of agent interactions impose strict conditions on existence the set of possible solutions. But it has to be seen if, for any arbitrary G graph (including or not including loops made by an odd number of agents), the set of non-self-consistent solutions described by Eq. (10) (Fig. 3(a, d)) is evolutionary stable. Despite such solution is mathematically possible, it might be highly unstable. If this intuition is proved, it would reinforce the idea that the lexical matrix has to be understood as a result of evolution, not as a primary feature of the internal endowment of agents.

The authors thank the members of the Complex Systems Lab and an anonymous referee for useful comments. This work has been supported by grants FIS2004-0542, IST-FET ECAGENTS project of the European Community founded under EU R&D contract 011940 and by the Santa Fe Institute.

References

Bollobás, B., 1998. Modern Graph Theory. Springer, New York.
Dorogovtsev, S., Mendes, J.F.F., 2002. Evolution of Networks. Oxford University Press, Oxford.
Ferrer-Cancho, R., Solé, R.V., 2003. Least effort and the origins of scaling in human language. Proc. Natl Acad. Sci. U.S.A. 100, 788–791.
Ferrer-Cancho, R., Bollobas, B., Riordan, O., 2005. The consequences of Zipf's law for syntax and symbolic reference. Proc. Roy. Soc. Lond. B 272, 561–565.
Hashimoto, T., 1997. Development of syntactic and semantic structure in artificial agents. In: Wang, PP. (Ed.), Joint Conference of Information Sciences, vol. 3. Rough Set and Computer Science. Duke University, Durham, pp. 238–241.
Hauser, M.D., Chomsky, N., Fitch, W.T., 2002. The faculty of language: what is it, who has it, and how did it evolve?. Science 298, 1569–1579.
Komarova, N.L., Niyogi, P., 2004. Optimizing intelligibility of linguistic agents in a shared world. Artif. Intell. 154, 1–42.
Komarova, N.L., Nowak, M.A., 2001. The evolutionary dynamics of the lexical matrix. Bull. Math. Biol. 63 (3), 451–485.
Lusseau, D., 2003. The emergent properties of a dolphin social network. Proc. Roy. Soc. Lond. B 271 (Suppl.), 5477–5481.
Newman, M.E.J., 2003. The structure and function of complex networks. SIAM Rev. 45, 167–256.
Nowak, M.A., Krakauer, D.C., 1999. The evolution of language. Proc. Natl Acad. Sci. U.S.A. 96, 8028–8033.
Nowak, M.A., Plotkin, J.B., Krakauer, D.C., 1999. The evolutionary language game. J. Theor. Biol. 200, 147–162.
Solé, R.V., 2005. Syntax for free?. Nature 434, 289.
Solé, R.V., Ferrer Cancho, R., Montoyo, J.M, Velverde, S., 2002. Selection, tinkering and emergence in complex networks. Complexity 8, 20–33.
Steels, L., 2000. The emergence of grammar in communicating autonomous robotic agents. In: Horn, W. (Ed.), Proceedings of ECAI 2000. IOS Publishing, Amsterdam.
Steels, L., 2003. Evolving grounded communication for robots. Trends Cognitive Sci. 7 (7), 308–312.
Watts, D.J., Strogatz, S.H., 1998. Collective dynamics of small-world networks. Nature 393, 440–442.
Wilson, R.J., 1996. Introduction to Graph Theory. Addison-Wesley, Essex, UK.

Bernat Corominas*, Ricard V. Solé**
ICREA-Complex Systems Lab, Universitat Pompeu Fabra, Dr. Aiguader 80, 08003 Barcelona, Spain
E-mail addresses: bernat.corominas@upf.edu (B. Corominas)

Ricard V. Solé**
Santa Fe Institute, 1399 Hyde Park Road, New Mexico 87501, USA
E-mail address: ricard.sole@upf.edu

*Corresponding author. Tel.: +34 935422834.
**Also for correspondence.