# REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

| 1. REPORT DATE (DD-MM-YYYY) | 2. REPORT TYPE | 3. DATES COVERED (From – To) |
|---|---|---|
| 02-10-2006 | Final Report | 3 June 2005 - 05-Dec-06 |

**4. TITLE AND SUBTITLE**

From Cognition To Language: The Modeling Field Theory Approach

**5a. CONTRACT NUMBER**
FA8655-05-1-3031

**5b. GRANT NUMBER**

**5c. PROGRAM ELEMENT NUMBER**

**6. AUTHOR(S)**

Professor Jose F Fontanari

**5d. PROJECT NUMBER**

**5d. TASK NUMBER**

**5e. WORK UNIT NUMBER**

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**
Universidade de Sao Paulo
Av. Trabalhador Sao Carlense 400
Caixa Postal 369
Sao Carlos CEP 13560-970
Brazil

**8. PERFORMING ORGANIZATION REPORT NUMBER**

N/A

**9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)**

EOARD
PSC 821 BOX 14
FPO AE 09421-0014

**10. SPONSOR/MONITOR'S ACRONYM(S)**

**11. SPONSOR/MONITOR'S REPORT NUMBER(S)**
Grant 05-3031

**12. DISTRIBUTION/AVAILABILITY STATEMENT**

Approved for public release; distribution is unlimited.

**13. SUPPLEMENTARY NOTES**

**14. ABSTRACT**

This report results from a contract tasking Universidade de Sao Paulo as follows: The Grantee will investigate a theoretical possibility for the integration of Language and Cognition, aiming to show it is coherent and plausible. The main goal is to improve the general understanding of issues concerning cognition and language making use of minimal models. Prototypes of hardware or software are not included, although depending on the results of the research, that will be a natural next step. Specific goals are 1) To determine whether communication among software agents can help in the solution of the differentiation problem, i.e., the development of more detailed knowledge. To achieve this requires that agents be endowed with a mechanism to determine the real number of objects in their world as well as mapping between concepts and signals. And 2) To introduce a second-level of cognitive hierarchy in order to make agents capable of recognizing relationships among objects. At the linguistic level this problem is related to the evolution of syntactic communication. The specific problem to be studied is the enemy-pack problem (simultaneous recognition of several targets). The main expected product of this research is a better understanding of the trade-off between cognition and communication.

**15. SUBJECT TERMS**
EOARD, Modeling & Simulation, Cognition, Language processing, Software Agents, Language

| 16. SECURITY CLASSIFICATION OF: | | | 17. LIMITATION OF ABSTRACT | 18, NUMBER OF PAGES | 19a. NAME OF RESPONSIBLE PERSON |
|---|---|---|---|---|---|
| **a. REPORT** | **b. ABSTRACT** | **c. THIS PAGE** | UL | | PAUL LOSIEWICZ, Ph. D. |
| UNCLAS | UNCLAS | UNCLAS | | 48 | **19b. TELEPHONE NUMBER** (Include area code) +44 20 7514 4474 |

**Standard Form 298** (Rev. 8/98)
Prescribed by ANSI Std. Z39-18

# Final Report


**From cognition to language: the  modeling field theory approach**

**Grant FA8655-05-1-3031**

José Fernando Fontanari
Instituto de Física de São Carlos, Universidade de São Paulo,
Caixa Postal 369
13560-970 São Carlos SP  Brazil
Tel: +55-16-33739849, Fax: +55-16-33739877, e-mail: fontanari@ifsc.usp.br

# Contents

# 1 Research activities

The main results of the research activities supported by EOARD were described in great detail and made public in the seven papers listed below. This was a very productive period in which several aspects of language evolution and meaning creation were investigated, as can be appreciated by the wide scope of the topics addressed in the publications.

1. José F. Fontanari and Leonid I. Perlovsky, *"Evolving compositionality in evolutionary language games"*, submitted to IEEE Transactions on Evolutionary Computation.
2. José F. Fontanari and Leonid I. Perlovsky, *"Allee effect on language evolution"* in The Evolution of Language, Proceedings of the 6th International Conference, edited by A. Cangelosi, A. D. M. Smith, K. Smith, World Scientific, Singapore, 2006, pp. 411-412.
3. Leonid I. Perlovsky and José F. Fontanari, *"How language can guide intelligence"* in The Evolution of Language, Proceedings of the 6th International Conference, edited by A. Cangelosi, A. D. M. Smith, K. Smith, World Scientific, Singapore, 2006, pp. 438-439.
4. José F. Fontanari and Leonid I. Perlovsky, *"Meaning creation and communication in a community of agents"*, to appear in the Proceedings of the International Joint Conference on Neural Networks (IJCNN06), Vancouver, Canada.
5. José F. Fontanari and Leonid I. Perlovsky, *"Categorization and symbol grounding in a complex environment"*, to appear in the Proceedings of the International Joint Conference on Neural Networks (IJCNN06), Vancouver, Canada.
6. Vadim Tikhanoff, José F. Fontanari, Angelo Cangelosi, and Leonid I. Perlovsky. *"Language and Cognition Integration through Modeling Field Theory: Category Formation for Symbol Grounding"*, to appear in the proceedings of the 16th International Conference on Artificial Neural Networks, ICANN 2006, Athens, Greece. The conference proceedings will be published in the Springer-Verlag series "Lecture Notes in Computer Science".
7. José F. Fontanari, *"Statistical analysis of discrimination games"*, submitted to Physical Review E.

These papers, which address the main topics of investigation of the original research proposal, are appended to the end of this report (see contents). Of particular relevance for the continuity of this research effort is the collaboration with the group lead by Dr. Cangelosi at University of Plymouth, that should focus on the extension of the results reported in extended abstract "How language can guide intelligence", item 3 of the above publication list. In fact, the finding that the exchange of information between two MFT categorization systems (or agents) can greatly improve the discriminating capability of each agent may be of some practical use. However, more research is needed since many crucial issues remain unexplored. For instance, in the present formulation, the agents observe distinct set of objects (or distinct parts of the environment) and, after categorization, exchange the labels (names) of the objects they discriminated. In this communication

stage, each agent observes the entire environment. What happens if in the previous stage the agents' observation set overlap, so they give different names (labels) to the same object? Or, if what the agents observe are different parts of the same object, would they, after communication, realize they are facing only one instead of two objects? These exciting questions will be tackled in the near future, so partial results can be published in the Proceedings of the KIMAS07.

## 2  Use of the award resources

As pointed out in the previous report, the funds corresponding to the first part of the EOARD award ($ 4,000.00) were used to cover part of the travel expenses to participate of the Evolang06 in Rome and of a work meeting in João Pessoa, Brazil. The second payment ($ 8,000.00) has been just incorporated into the budget of the Instituto de Física de São Carlos (IFSC) and will be partially used to support my participation in the IJCNN06 in Vancouver, Canada as well as in KIMAS07 in Boston, USA. We advance that the funds of the final payment ($ 10,000.00) will be used to strengthen the collaboration with the researchers at the University of Plymouth.

## 3 Acknowledgement of Sponsorship

## 4 Disclaimer

The views and conclusions contained herein are those of the author and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Air Force Office of Scientific Research or the U.S. Government.

## 5 Disclosure of inventions

I certify that there were no subject inventions to declare during the performance of this grant.

# Evolving compositionality in evolutionary language games

José F. Fontanari, and Leonid I. Perlovsky, *Member, IEEE*

*Abstract*— **Evolutionary language games have proved a useful tool to study the evolution of communication codes in communities of agents that interact among themselves by transmitting and interpreting a fixed repertoire of signals. Most studies have focused on the emergence of Saussurean codes (i.e., codes characterized by an arbitrary one-to-one correspondence between meanings and signals). In this contribution we argue that the standard evolutionary language game framework cannot explain the emergence of compositional codes - communication codes that preserve neighborhood relationships by mapping similar signals into similar meanings – even though use of those codes would result in a much higher payoff in the case that signals are noisy. We introduce an alternative evolutionary setting in which the meanings are assimilated sequentially and show that the gradual building of the meaning-signal mapping leads to the emergence of mappings with the desired compositional property.**

*Index Terms*— **Complexity Theory, Game theory, Genetic algorithms, Simulation**

## I. INTRODUCTION

THE case for the study of the evolution of communication within a multi-agent framework was probably best made by Ferdinand de Saussure in his famous statement "language is not complete in any speaker; it exists only within a collectivity... only by virtue of a sort of contract signed by members of a community" [1]. Translated into the biological jargon, this assertion means that language is not the property of an individual, but the extended phenotype of a population [2]. More than one decade ago, seminal computer simulations were carried out to demonstrate that cultural [3] as well as genetic [4] evolution could lead to the emergence of ideal communication codes (i.e., arbitrary one-to-one correspondences between objects or meanings and signals), termed Saussurean codes, in a population of interacting agents. Typically, the behavior pattern of the agents was modeled by (probabilistic) finite state machines. The work by Hurford [3], in particular, set the basis of the Iterated Learning Model (ILM) for the cultural evolution of language, the typical realization of which consists of the interaction between two agents - a pupil that learns the language from a teacher

[5]. In those studies, language is viewed as a mapping between meanings and signals. The communication codes that emerged from the agents interactions are, in general, non-compositional or holistic communication codes, in which a signal stands for the meaning as a whole. In contrast, a compositional language is a mapping that preserves neighborhood relationships – similar signals are mapped into similar meanings. If there is a nontrivial structure in both meaning and signal spaces then, in certain circumstances, making explicit use of those structures may greatly improve the communication accuracy of the agents. The emergence of compositional languages in the ILM framework beginning from holistic ones in the presence of bottlenecks on cultural transmission was considered a breakthrough in the computational language evolution field [5]-[7]. The aim of this contribution is to understand how compositional communication codes can emerge in an evolutionary language game framework [3],[4],[8],[9].

The way we introduce the structure of the signal space (i.e., the notion of similarity between signals) into the rules of the language game is through errors in perception: the signals are assumed to be corrupted by noise so that they can be mistaken for one of their neighbors in signal space [8]. Similarly, the structure of the meaning space enters the game by rewarding the agents that, prompted by a signal, infer a meaning close to the meaning actually intended by the emitter. Of course, the reward for incorrect but close inferences must be smaller than that granted for the correct inference of the intended meaning (see [9] for a similar approach). Hence the role played by noise in this context is similar to the role of the bottleneck transmissions in the ILM framework, since both make advantageous the exploration of the detailed structure of the meaning-signal mapping. In particular, we show that once a Saussurean communication code is established in the population, i.e., all agents use the same code, it is impossible for a mutant to invade, even if the mutant uses a better code, say, a compositional one. This is essentially the Allee effect [10], [11] of population dynamics which asserts that intraspecific cooperation might lead to an inverse density dependence, resulting in the extinction of some (social) animal species when their population size becomes small. Of course, this effect is germane to the outcome of biological

invasions involving such species. We note that most realizations of the ILM circumvent this difficulty by assuming that the population is composed of two agents only, the teacher and the pupil, and that the latter always replaces the former. However, according to de Saussure (see quotation above), this is not an acceptable framework for language. In addition, a bias toward compositionality is built in the inference procedure used by the pupil to fill in the gaps due to transmission bottlenecks, in which some of the meanings are not taught to the pupil. This bias towards generalization, together with cultural evolution, seems to be the key ingredients to evolve compositionality in the ILM framework. Understanding as well as demonstrating how innovations that increase the expressive power of individuals can spread through a population is the essence of any evolutionary explanation to language evolution [9]. Accordingly, the solution we propose to the problem of evolving a compositional code in a population of agents that exchange signals with each other and receive rewards at every successful communication event is the incremental assimilation of meanings, i.e., the agents construct their communication codes gradually, by seeking a consensus signal for a single meaning at a given moment. Only after a consensus is reached, a novel meaning is permitted to enter the game. This sequential procedure, which dovetails with the classic Darwinian explanation to the evolution of strongly coordinated system, allows for the emergence of fully compositional codes, an outcome that we argue is very unlikely, if not impossible, in the traditional language game scenario in which the consensus signals are sought simultaneously for the entire repertoire of meanings.

## II. MODEL

Here we take the more conservative viewpoint that language evolved from animal communication as a means of exchanging relevant information between individuals rather than as a byproduct of animal cognition or representation systems (see, e.g., [12], [13] for the opposite viewpoint). In particular, we consider a population composed of $N$ agents who make use of a repertoire of $m$ signals to exchange information about $n$ objects. Actually, since the groundbreaking work of de Saussure [1] it is known that signals refer to real-world objects only indirectly as first the sense perceptions are mapped onto a conceptual representation – the meaning – and then this conceptual representation is mapped onto a linguistic representation – the signal. Here we simply ignore the object-meaning mapping (see, however, [14], [15]) and use the words object and meaning interchangeably. To model the interaction between the agents we borrow the language game framework proposed by Hurford [3] (see also [8]) and assume that each agent is endowed with separate mechanisms for transmission (i.e., communication) and for reception (i.e., interpretation). More pointedly, for each agent we define a $n \times m$ transmission matrix $P$ whose entries $p_{ij}$ yield the probability that object

$i$ is associated with signal $j$, and a $m \times n$ reception matrix $Q$ the entries of which, $q_{ji}$, denote the probability that signal $j$ is interpreted as object $i$. Henceforth we refer to $P$ and $Q$ as the language matrices. In general, the entries of these two matrices can take on any value in the range $[0,1]$ satisfying the constraints $\sum_{j=1}^{m} p_{ij} = 1$ and $\sum_{i=1}^{n} q_{ji} = 1$, in conformity with their probabilistic interpretation. In this contribution, however, we consider the case of binary matrices, in which the entries of $Q$ and $P$ can assume the values 0 and 1 only. There are two reasons for that. First, in the absence of errors in language learning, the evolutionary language game will eventually lead to binary transmission and reception matrices, regardless of the values of $m$ and $n$, and of the initial choice for the entries of those matrices [16]. So our restriction of the entry values to binary quantities has no effect on the equilibrium solutions of the evolutionary game. In addition, these deterministic encoders and decoders were shown to always perform better than their stochastic variants [17]. Second, by assuming that the transmission and reception matrices are binary we recover the synthetic ethology framework proposed by MacLennan [4], a seminal agent-based work on the evolution of communication in a population of finite state machines (see also [18]).

Although the reception matrix $Q$ is, in principle, independent of the transmission matrix $P$, results of early computer simulations have shown that, in a noiseless environment, the optimal communication strategy is the Saussurean two-way arbitrary relationship between an object and a signal, i.e., the matrices $P$ and $Q$ are linked such that if $p_{ij} = 1$ for some object-signal pair $i, j$ then $q_{ji} = 1$ [3]. These matrices are associated to the Saussurean communication codes introduced before, provided there are no correlations between the different rows of the matrix $P$, i.e., the assignment object-signal is arbitrary.

### A. The evolutionary language game

Given the transmission and reception matrices, the communicative accuracy or overall payoff for communication between two agents, say $I$ and $J$, is defined as [3],[8],[19]

$$F(I,J) = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{m} \left( p_{ij}^{(I)} q_{ji}^{(J)} + p_{ij}^{(J)} q_{ji}^{(I)} \right) \qquad (1)$$

from which we can observe the symmetry of the language game, i.e., both signaler and receiver are rewarded whenever a successful communication event takes place. By assuming such a symmetry, one ignores a serious hindrance to the evolution of language: passing useful information to another agent is an altruistic behavior [20], [21] that can be maintained in human societies thanks to the development of reciprocal altruism, in which unrelated individuals mutually benefit by exchanging the donor and the receiver roles multiple times [22]. However, the scarcity of empirical demonstrations of reciprocal altruism in nature, except for modern humans, motivated an alternative scenario for the evolution of

language, namely, that human language evolved as a "mother tongue" – a communication system used among kin, especially between parents and their offspring [23].

In this contribution, we assume the validity of Eq. (1) and simply ignore the costs of honest signalling [20]. Hence we take for granted the existence of special social conditions to foster reciprocal altruism among the agents or, alternatively, a mother tongue scenario in which the agents are related to each other. In this vein, it is interesting to note that although in the work by MacLennan [3] communication is defined following Burghardt [24] as "the phenomenon of one organism producing a signal that, when responded to by another organism, confers some advantage to the signaler or his group" (see [25] for alternative definitions of communication), the actual implementation of the simulation rewards equally the two agents that take part in the successful communication event. In the case where only the receiver is rewarded, Saussurean communication fails to evolve [26].

Assuming, in addition, that each agent $I$ interacts with every other agent $J = 1, \ldots, N$ $(J \neq I)$ in the population we can immediately write down the total payoff received by $I$,

$$F_I = \frac{1}{N-1} \sum_{J \neq I} F(I, J) \,, \tag{2}$$

in which the sole purpose of the normalization factor is to eliminate the trivial dependence of the payoff measure on the population size $N$. Following the basic assumption of evolutionary game theory [27] this quantity is interpreted as the fitness of agent $I$. Explicitly, we assume that the probability that $I$ contributes with an offspring to the next generation is given by the relative fitness

$$w_I = F_I \Big/ \sum_J F_J \,, \tag{3}$$

which essentially implies that mastery of a public communication system adds to the reproductive potential of the agents [3].

There are several distinct ways to implement the language game. For instance, MacLennan [4] and Fontanari & Perlovsky [18] stick to the genetic algorithm approach (see, e.g., [28]) in which the offspring acquires both the transmission and reception matrices from its parent, assuming clonal or asexual reproduction. The offspring is identical to its parent except for the possibility of mutations that may alter a few rows of the language matrices. However, here we take a different viewpoint and reinterpret this genetic model within a learning context. We assume, in particular, that the offspring actually learns the language from its parent but that the learning is not perfect – there is a probability $\mu$ that the communication code it acquires is slightly different from its parent's. This very framework has been used to study the emergence of universal grammar and syntax in language [2],[29], [30].

An alternative learning scenario used by Nowak & Krakauer [8] assumes that the offspring adopt the language of its parent by sampling its response to every object $k$ times. This approach makes sense only if the language matrices are not binary, though, as mentioned before, in the long run those matrices must become binary. For $k \to \infty$, the offspring is identical to its parent, which corresponds then to $\mu = 0$ in the previous learning scenario, whereas differences between parent and offspring arise in the case of finite $k > 1$. This sampling effect is qualitatively similar to the effect of learning errors in the scenario introduced before. For $k = 1$, already the first generation of offspring becomes represented by binary language matrices and so the sampling procedure is rendered ineffective. The reason is that a binary matrix P assigns each object to a unique signal (though this same signal can be used also for a distinct object), and so sampling the responses of the parent to the same object will always yield the same signal. As a result, the evolutionary process based on learning by sampling halts - the offspring become identical to their parents.

A similar but more culturally inclined approach is that followed by Hurford [3] and Nowak et al. [16]: instead of sampling the parent's responses, the offspring samples the responses of a certain number of agents in the population or even of the entire population. In this case, the hereditary component is lost since the offspring, in general, will not resemble its parent, and so evolution by natural selection has no say in the outcome of the dynamics. In the case of Hurford [3] there is still a strong genetic component as the offspring inherits from its parent its strategy of inference. Similarly, the Iterated Learning Model (ILM) for the cultural evolution of language (see [5], [7] for reviews) in its more popular version consists of two agents only, the teacher and the pupil who learns from the teacher through a sampling process identical to that just described. The pupil then replaces the teacher and a new, tabula rasa pupil is introduced in the scenario. This procedure is iterated until convergence is achieved. In this case, the payoff (2) plays no role at all in the language evolutionary process and the stationary language matrices will depend strongly on the inference procedure used by the pupil to create a meaning/signal mapping from the teacher responses. Of particular interest for our purpose is the finding that compositional codes emerge in the case that the learning strategy adopted by the pupil supports generalization and that this ability is favored by the introduction of transmission bottlenecks in the communication between teacher and pupil. Such a bottleneck occurs when the learner does not observe the signal for some objects. This contrasts with the sampling effect mentioned before in which the learner observes the signals to every object. In this contribution we study whether and in what conditions compositional codes emerge in an evolutionary language game.

*B. The meaning-signal mapping*

As already pointed out, language is viewed as a mapping between objects (or meanings) and signals and compositionality is a property of this mapping: a compositional language is a mapping that preserves neighborhood relationships, i.e., nearby meanings in the meaning space are likely to be associated to nearby signals in

signal space [5]. At first sight, this notion looks contradictory to the well-established fact that the relation between a word (signal) and its meaning is utterly arbitrary. For instance, as pointed out by Pinker [31], "babies should not, and apparently do not, expect *cattle* to mean something similar to *battle*, or *singing* to be like *stinging*, or *coats* to resemble *goats*". In fact, Pettito demonstrated that the arbitrariness of the relation between a sign and its meaning is deeply entrenched in the child's mind [32]. On the other hand, sentences like *John walked* and *Mary walked* have parts of their semantic representation in common (someone performed the same act in the past) and so the meaning of these sentences must be close in the meaning space. Since both sentences contain the word *walked* they must necessarily be close in signal space as well. Following Pinker, we acknowledge a significant degree of arbitrariness at the level of word-object pairing. This might be a consequence of a much earlier (pre-human) origin of this mechanism, as compared to seemingly distinctly human mind mechanisms for sentence-situation pairing. From a mathematical modeling perspective, however, such a distinction is not essential for our purposes, since the signals (sentences or words) can always be represented by a single symbol - only the "distance" between them will reflect the complex inner structure of the signal space. For instance, suppose there are only two words which we represent, without lack of generality, by 0 and 1 so that any sentence could be described as a binary sequence and so represented by a single integer number. Here the relevant distance between two such sentences is the Hamming distance rather than the result of the subtraction between their labeling integers. This notion, of course, generalizes trivially to the case when the sentences are composed of more than two types of words.

For simplicity, in this contribution we consider the case where both signals and meanings are represented by integer numbers and the relevant distance in both signal and meaning space is the result of the usual subtraction between integers. Figure 1 illustrates one of the $n \times m$ possible meaning-signal mappings. A quantitative measure of the compositionality of a communication code is given by the degree to which the distances between all the possible pairs of meanings correlates with the distance between their corresponding pairs of signals [7]. Explicitly, let $\Delta m_{ij}$ be the distance between meanings $i$ and $j$, and $\Delta s_{ij}$ the distance between the signals associated to these two meanings. Introducing the averages $\Delta \overline{m} = \sum_{(ij)} \Delta m_{ij} / p$ and $\Delta \overline{s} = \sum_{(ij)} \Delta s_{ij} / p$ where the sum is over all distinct pairs $p = n(n-1)/2$ of meanings, the compositionality of a code is defined as the Pearson correlation coefficient [7]

$$C = \frac{\sum_{(ij)} \left( \Delta m_{ij} - \Delta \overline{m} \right) \left( \Delta s_{ij} - \Delta \overline{s} \right)}{\left[ \sum_{(ij)} \left( \Delta m_{ij} - \Delta \overline{m} \right)^2 \sum_{(ij)} \left( \Delta s_{ij} - \Delta \overline{s} \right)^2 \right]^{1/2}} \qquad (4)$$

so that $C \approx 1$ indicates a compositional code and $C \approx 0$ an unstructured or holistic code. This definition applies only to codes that implement a (not necessarily arbitrary) one-to-one correspondence between meaning and signal.
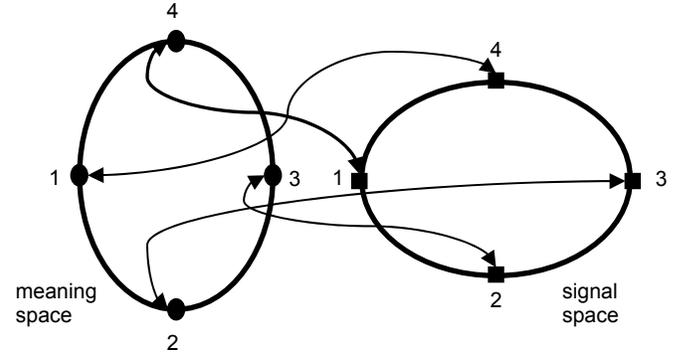


Figure 1: Example of a mapping meaning-signal for $n = m = 4$. The integers here may be viewed as labels for complex entities (e.g., sentences). The large circles indicate cyclic boundary conditions so that, e.g., signal 1 is 1 unit distant from signals 2 and 4. The code represented in the Figure has compositionality $C = 1$.

Strictly, here we do not address directly the emergence of compositionality, defined as the property that the meaning of a complex expression is determined by the meanings of its parts and the rules used to combine them. Rather, we focus on the emergence of structured communication codes, which preserve the topology of the meaning-signal mapping, in that similar meanings are associated with similar signals and vice-versa. It seems that an important aspect of joint evolution of compositional cognition and compositional language is their evolution along with structural metric (or approximately metric) spaces of cognition and meaning. In this contribution we assume that a metric space exists, and explore the consequences for the emergence of compositionality. The connection between structured and compositional meaning-signal mappings can be made explicit if we consider an artificial scenario for which there is a prescription to derive the meaning of the whole given the meaning of the elementary parts. (Such prescription is clearly ruled out in real language since context and previous knowledge play a crucial role in our understanding of any situation.) In this case the distance between any two complex meanings could be inferred by comparing their components and, consequently, by introducing a metric in the meaning space.

Our approach ties in with the view that properties of language such as compositionality are emergent characteristics of the explosion of semantic complexity occurred during hominid evolution [33]. Semantic complexity means not only a large number of cognitive categories (meanings) but also an increase in their perceived interrelationships, which are inherent properties of the topology of the meaning space. In fact, the number of objects for which a person has separate words is not too large: a recent estimate suggests a vocabulary of around 17,000 base words for well-educated adult native speakers of English [34]. This is a not a very big number and so it is reasonable to assume that object-word associations can be learned from examples, one by one. The number of situations which are combinations of objects, on the other hand, is larger than the number of all elementary particle events in the history of the Universe. This supports a need for the assumption of compositionality in language. As

hinted in [33], a natural avenue to study the evolution of complex features of language (e.g., compositionality) is the increase of the complexity of the meaning space, which is exactly the approach offered in this contribution.

*C. Errors in perception*

So far as the communicative accuracy introduced in Eq. (1) is concerned, the structures of the meaning and signal spaces are irrelevant to the outcome of the evolutionary language game: the total population payoff is maximized when all agents adopt a code that implements a one-to-one correspondence between meanings and signals. Such a code is, of course, described by any one of the $n!$ permutation language matrices. The fact that ultimately all agents adopt the same communication code is a general result of population genetics related to the effect of genetic drift on a finite population [35]. To permit that the structure of the meaning and signal spaces play a role in the evolutionary game and so to break the symmetry among the permutation matrices so as to favor the compositional codes we must introduce a new ingredient in the language game, namely, the possibility of errors in perception [8]. In fact, it is reasonable to assume that in the earlier stages of the evolution of communication the signals were likely to be noisy and so they could be easily mistaken for each other. The relevance of the structure of the signal space becomes apparent when we note that the closer two signals are, the higher the chances that they are mistaken for each other. This aspect of the model can be described by an agent-independent $m \times m$ confusion matrix $E$, the entries of which $e_{ij}$ yield the probability of signal $j$ being observed as signal $i$ due to corruption by noise [8],[9].

To introduce the structure of the meaning space in the language game, we note first that Eq. (1) has a simple interpretation in the case of binary, but not necessarily permutation, language matrices: both signaler and receiver are rewarded with $\frac{1}{2}$ unity of payoff whenever the receiver interprets correctly the meaning of the emitted signal. Otherwise, there is no reward to any of the two parts, no matter how close the inferred meaning is from the correct one. This gives us a clue of how to modify the model in order to take into account the meaning structure – just ascribe some small reward value to both agents if the inferred meaning is close to the intended one. In fact, giving value to decisions which are not the best ones is a common assumption in decision and game theory [36] and seems to be consistent with what is actually observed in nature since, clearly, not every misinterpretation is equally harmful [9]. Consider for instance the Vervet monkey alarm calls [37]: misinterpreting a snake alarm for a leopard one, and hence running to a tree instead of standing up and looking in the grass, is clearly much better than misinterpreting it for an eagle call.

Following Nowak et al. [8] and Zuidema [9], we can formalize the notion of meaning similarity by introducing another agent-independent matrix, the $n \times n$ value matrix $V$, so that $v_{ij}$ yields the payoff attributed to an agent which infer meaning $i$ when the actual meaning the signaler intended to

transmit was $j$. Hence the overall payoff for communication between agents $I$ and $J$, becomes [9]

$$F(I,J) = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} v_{ij} \left[ P^{(I)} \times \left( E \times Q^{(J)} \right) + P^{(J)} \times \left( E \times Q^{(I)} \right) \right]_{ij}$$

(5)

where $\times$ stands for the usual matrix multiplication. Note that Eq. (1) is recovered in the case that both value and confusion matrices are diagonal.

In particular, here we will consider the simple case in which there is a probability $\varepsilon \in [0,1]$ that a signal be mistaken for one of its nearest neighbors, i.e., $e_{i,j} = \varepsilon/2 \left( \delta_{i,j+1} + \delta_{i,j-1} \right)$. So, in the example of Fig. 1, signal 4 can be mistaken only for signals 3 or 1 (remember the cyclic structure of the signal space) with probability $\varepsilon$. Similarly, agents are rewarded only if the inferred meaning is one of the nearest neighbors of the intended meaning, i.e. $v_{ij} = r \left( \delta_{i,j+1} + \delta_{i,j-1} \right)$, or, of course, the intended one $v_{ii} = 1$. Here $r \in [0,1]$ is a parameter that measures the advantage, in terms of payoff, of using a compositional communication code rather than a Saussurean one.

Together with the presence of noise, this last ingredient – nonzero reward for inferring a meaning close to the correct one – should, in principle, favor the emergence of compositional communication codes in an evolutionary game guided by Darwinian rules. In what follows we will show that the problem of evolving efficient communication codes within an evolutionary framework, whether in the presence or not of noise, is more difficult than previously realized [4], [16], [18]. This problem differs from usual optimization problems tackled with evolutionary algorithms in that the maximization of the average population payoff requires a somewhat coordinated action of the agents. It is of no value for an agent to exhibit the correct "genome" (i.e., the transmission and reception matrices) if it cannot communicate efficiently with the other agents in the population because they use different language matrices.

The emergent view of compositionality adopted here differs from the approach followed by Nowak et al [29] to study the evolution of syntactic (or combinatorial) communication. In that work the conditions at which syntax are advantageous over non-syntactic or holistic languages were determined, namely, when the number of required signals to express the relevant meanings exceeds some threshold value. (It should be noted that combinatorial communication has its disadvantages too, since it boosts the potential for deception [38].) However, the finding that the adoption of a particular communication code is better for the population, in that it yields an higher overall payoff, is no guarantee that such code will actually spread in the population. On the contrary, in this contribution we show that the Allee effect will prevent its spreading. Additional assumptions, such as the semantic continuity of incremental learning proposed here, seem to be necessary to guarantee the emergence of compositional codes.

### III. POPULATION DYNAMICS

We assume that the offspring learn their languages from their parents. Were it not for the effect of errors during learning, which results in small changes in the language matrices, the offspring would be identical to their parents. Like mutations in the genetic setup, these learning errors allow for the variability of the agents, and thus for the action of natural selection.

We start with $N$ agents (typically $N = 100$) whose binary language matrices are set randomly. Explicitly, for each agent and for each meaning $i = 1, \ldots, n$ we choose randomly an integer $j \in \{1, \ldots, m\}$ and set $p_{ij} = 1$ and $p_{ik} = 0$ for $k \neq j$. Similarly, for each signal $j = 1, \ldots, m$ we choose an integer $i \in \{1, \ldots, n\}$ and set $q_{ji} = 1$ and $q_{jk} = 0$ for $k \neq i$. This procedure guarantees that initially $P$ and $Q$ are independent random probability matrices. Note that, in general, they are not permutation matrices at this stage. To calculate the total payoff of a given agent, say agent $I$, we let it to interact with every other agent in the population. At each interaction, the emitted signal can be mistaken for one of the neighboring signals with probability $\varepsilon$. According to Eq. (5), at each communication event (an interaction) agent $I$ receives the payoff value $\frac{1}{2}$ if the receiver guesses the intended meaning of the signal that $I$ has emitted, the payoff value $\frac{r}{2}$ if the receiver guessing is one of the nearest neighbors of the intended meaning, and payoff value 0 otherwise. Of course, the receiver obtains the same payoff accrued to agent $I$. Once the payoffs or fitness of all $N$ agents are tabulated, the relative payoffs can be calculated according to Eq. (3), and then used to select the agent that will contribute with one offspring to the next generation.

To keep the population size constant we must eliminate one agent from the population. To do that we will use two strategies: (i) to choose the agent to be eliminated at random, regardless of its fitness value, and (ii) to use an elitist strategy which eliminates the agent with the lowest fitness value. In both cases, the recently produced offspring is spared from demise. The first selection procedure is Moran's model of population genetics [35]. Both procedures differ from the standard genetic algorithm implementation [28] in that they allow for the overlapping of generations, a crucial prerequisite for cultural evolution which may be relevant when learning is allowed. In practice, however, Moran's model does not differ from the parallel implementation in which the entire generation of parents is replaced by that of the offspring in a single generation. Finally, to allow for the appearance of novel codes (or language matrices) in the population, changes are performed independently on the transmission and reception matrices of the offspring with probability $u \in [0,1]$. Explicitly, the transmission matrix $P$ is modified by changing randomly the signal associated to an also randomly chosen meaning with probability $u$. A similar procedure updates the reception matrix $Q$. Hence the probability that the same offspring has its transmission and reception matrices

simultaneously altered by errors is $u^2$ and the probability that it will differ somehow from its parent is $\mu = 1 - (1 - u)^2$. Henceforth we will refer to $\mu$ as the probability of error in language acquisition.

To facilitate comparison between different evolutionary algorithms we define a properly normalized average payoff of the population

$$G = \frac{1}{nN} \sum_{I=1}^{N} F_I \, , \tag{6}$$

so that $G \in [0,1]$. The maximum value $G = 1$ is reached for Saussurean codes in the case of noiseless communication. In addition, we define the generation time $t$ as the number of generations needed to produce $N$ offspring with the consequent elimination of the same number of agents.

In Figure 2 we present the effect of the inaccuracy in language acquisition on the average payoff of the population for the simplest situation, namely, $\varepsilon = 0$ (the receiver always gets the original signal) and $r = 0$ (only inference of the correct meaning is rewarded). The results show a stark difference between the elitist and the usual evolutionary strategy regarding the form they are affected by learning errors. Whereas the performance of Moran's model is degraded for high error rates [39], reaching the payoff of random binary matrices for $\mu = 1$, the elitist strategy actually benefits from those errors and gets to the maximum payoff for the highest possible error rate. In fact, for small but nonzero values of the error rate, the communication accuracy of the elitist strategy is practically constant and starts to increases only after $\mu$ crosses some threshold value, $\mu \approx 0.02$. The performance of Moran's model, on the other hand, indicates the existence of an optimum value of the learning error for which the communication accuracy is maximum. Longer runs do not show any significant change of the pattern illustrated in Fig. 2. What enables the elitist strategy to take advantage of errors is the overlapping of generations together with the immediate removal of unfit agents from the population. This combination prevents the accumulation of those agents in the population and the consequent degradation of the communication performance observed in Moran's model. Moreover, by eliminating the agent that performs worse in the language game, the elitist strategy adds an extra kick to the selective pressure towards better communication codes, in addition to the offspring production regulated by the relative fitness, Eq. (3). Hence, in view of the remarkable effectiveness of the elitist strategy to maximize the communication accuracy of the population, in what follows we will present the results for that strategy only.
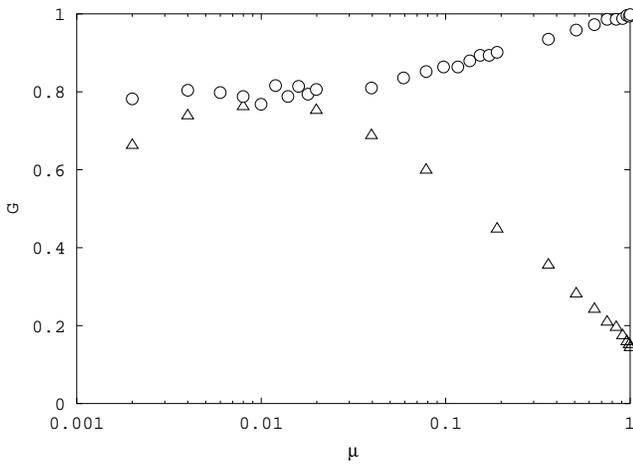
Figure 2: Normalized average payoff $G$ of the population as function of the probability of error in language acquisition $\mu$ in the case of $N=100$ agents communicating about $n=10$ meanings using $m=10$ signals. The evolution was followed until $t=2\times10^{3}$ for the elitist strategy ($\circ$) and until $t=10^{4}$ for Moran's model ($\Delta$). The symbols represent the average over 50 independent runs. The error bars are smaller than the symbol sizes. For $\mu=0$ we find $G=0.255\pm0.005$ for both strategies, whereas for random language matrices we find $G=0.1\pm0.0001$. The other parameters are $\varepsilon=r=0$. The search space is the $m^{n}\times n^{m}$ space spanned by the two independent binary probability matrices $P$ and $Q$.

Figure 3: Normalized average payoff for the elitist ($\circ$) strategy at $t=2\times10^{3}$ for 100 independent sample runs of the evolutionary dynamics. These results are compared to that of a fully compositional code (solid line) and of Saussurean codes ($\times$). The parameters and search space are the same as in Fig. 2 with $\mu=1$, except that now we have included a pressure for compositionality: the signals are corrupted with probability $\varepsilon=0.2$ and the ratio between the payoffs for inferring a close and the correct meaning is $r=0.25$. The optimal, compositional code yields $G\approx0.85$ and the typical payoff of a Saussurean code is $G\approx0.80$.

Figure 3 presents the average communication accuracy for 100 independent runs (populations) in a generic case in which the parameters $\varepsilon$ and $r$, which couple the dynamics with the distances in the signal and meaning spaces, are nonzero. Since now the communication between any two agents is affected by noise we must adopt a slightly different procedure to evaluate the payoff of the entire population. As before, we follow the evolutionary dynamics (i.e., the differential reproduction and learning-with-error procedures) until $t=2\times10^{3}$, then we store the language matrices of all $N$ agents. Keeping these matrices fixed we evaluate the average population payoff in 100 contests. A contest is defined by the interaction between all pairs of agents in the population. Actually, according to Eq. (5) each interaction comprises two communication attempts, since any given agent first plays the role of the emitter and then of the receptor. Hence a contest amounts to $N(N-1)$ communication events. Of course, in the noiseless case ($\varepsilon=0$) the payoff obtained would be the same in all contests. The procedural changes are needed to average out the effect of noise. For instance, in a single interaction two perfectly compositional codes could perform worse than two holistic codes if, by sheer chance, the signals happen to be corrupted only during the interaction of the compositional codes. To avoid such spurious effects the payoffs resulting from the interactions between any two agents are averaged out over 100 different interactions.
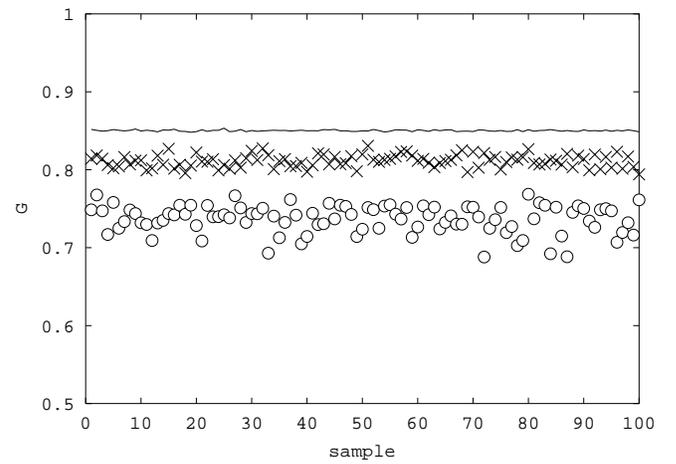
For the purpose of comparison, in Figure 3 we present also the results for a population of agents carrying the same perfectly compositional code ($C=1$) as well as for a similarly homogenous population of agents carrying identical Saussurean codes. These are control populations that, in contrast to the elitist populations, do not evolve. In the absence of noise, these control populations would reach the maximum allowed payoff, $G=1$. We note that a perfectly compositional code is not a Saussurean code, in the sense that the one-to-one mapping between meaning and signals is not arbitrary. The elitist strategy seems to face great difficulties even to find a Saussurean code, as compared with the performance in the noiseless case (see Figure 2) for instance, not to mention to find the optimum, perfect compositional code. Actually, in the presence of noise the performance of the Saussurean code seems to pose an upper limit to the performance of the elitist strategy by acting as an attractor to the evolutionary dynamics.

It is instructive to calculate the average payoff $G_{c}$ of a population composed of identical agents carrying a perfect compositional code. Consider the average payoff received by a given agent, say $I$, in a very large number of interactions with one of its siblings, say $J$. When $I$ plays the signaler role its average payoff is $(1-\varepsilon)\times1/2+\varepsilon\times r/2$, which, by symmetry, happens to be the same average payoff $I$ receives when it plays the receiver role. Since all agents are identical, the expected payoff of any agent equals that of the population. Hence

$$G_{c}=1-\varepsilon(1-r).\tag{7}$$

We can repeat this very same reasoning to derive the average payoff $G_s$ of a homogenous population of Saussurean codes. In this case, by playing the signaler, $I$ receives the average payoff $(1-\varepsilon) \times 1/2 + \varepsilon \times 2/(n-1) \times r/2$ where the factor $2/(n-1)$ accounts for the fact that the reward $r/2$ is obtained only if the inferred meaning is one of the two neighbors of the correct meaning. Hence this reasoning is valid for $n > 2$ only, since for $n = 2$ each meaning has a single neighbor, and so there is no difference between Saussurean and compositional codes. Taking into account the payoff received by $I$ when playing the receiver yields

$$G_s = 1 - \varepsilon + \frac{2\varepsilon}{n-1}r, \qquad (8)$$

for $n > 2$. Note that $G_c > G_S$ for $n > 3$. Similarly to the case $n = 2$, the Saussurean codes for $n = 3$ are compositional codes because of the cyclic boundary conditions in the meaning space. The values of the compositionality of the code carried by the agent with the largest payoff value in each of the runs are shown in Figure 4. Although there is a very slight tendency to compositionality in the codes produced by the elitist strategy, it is fair to say that the pressure to generate compositional code has not worked as expected, despite the clear advantage of such codes given the conditions of the experiment (see Figure 3). As pointed out, the reason for that might be that the Saussurean codes act as barriers (local maxima) from which the evolutionary dynamics cannot escape, thus impeding it from reaching a perfect compositional code (global maximum).
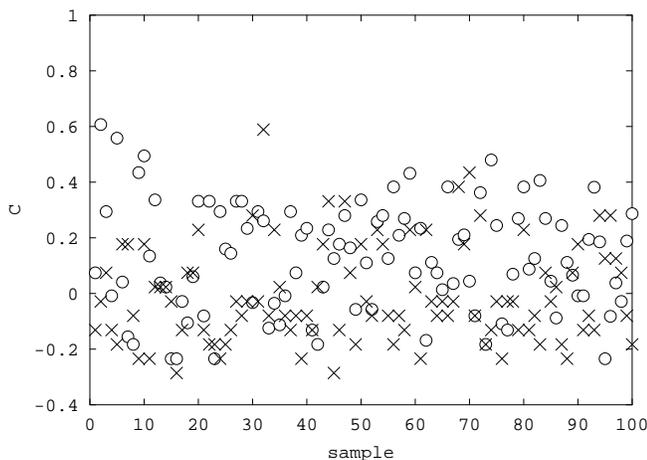


Figure 4: Compositionality of the code carried by the agent with the highest payoff in the runs shown in the previous Figure. The compositionality of the perfect compositional code is, by definition, $C = 1$. There is a slight tendency to compositionality in the codes produced by the elitist ($\circ$) strategy as compared to those of the Saussurean codes ($\times$).

The results depicted in Fig. 3 expose clearly the failure of the language evolutionary framework to produce efficient communication codes when the receiver must interpret noisy signals. To rule out the possibility that the cause of such failure was the initial unlikely decoupling between production and interpretation, in the following we will restrict the search space to that of Saussurean codes. Hence, for any agent, the transmission matrix $P$ is a permutation matrix and the reception matrix $Q$ has entries given by $q_{ji} = 1$ if $p_{ij} = 1$ and $0$ otherwise ($Q$ is also a permutation matrix). The initial population is composed of $N$ agents adopting distinct Saussurean codes. To guarantee that all new codes generated by mutations stay within our search space, we modify the mutation procedure so that with probability $\mu$ the signal associated to a randomly chosen meaning, say $i$, is exchanged with the signal associated to another randomly chosen meaning, say $k$. This corresponds to the interchange of the rows $i$ and $k$ of the transmission matrix. The reception matrix is then updated accordingly. The sole genetic strategy we use in the forthcoming simulations is the elitist one, in which the worst performing agent is replaced by the offspring of the agent chosen by rolling the fitness wheel.
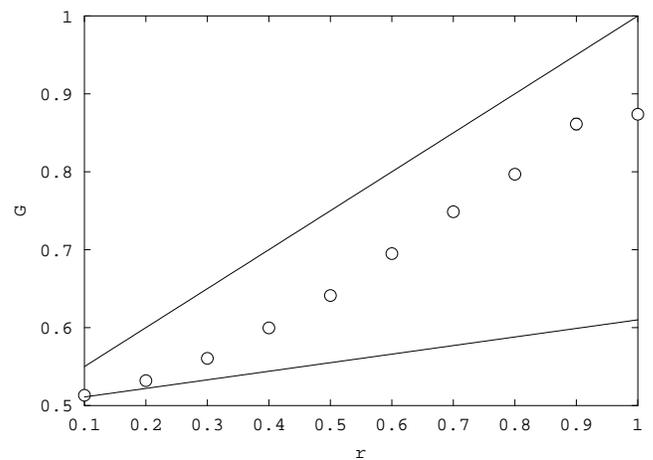


Figure 5: Average payoff resulting from 100 independent runs of the noisy evolutionary language game with the search space restricted to permutation matrices ($\circ$) as a function of the pressure for compositionality. The error bars are smaller than the symbol sizes. The upper straight line is the function $G_c = (1+r)/2$ that yields the average payoff of a perfect compositional code and the lower straight line is $G_S = 0.5 + 0.11r$ that yields the average payoff of a Saussurean code (see equations (7) and (8)). The parameters are $\varepsilon = 0.5$, $\mu = 0.9$, $N = 100$, and $n = m = 10$.

In Figure 5 we show the results of the experiments with the evolutionary search restricted to the space of permutation matrices. The procedure we use here was the same as that employed to draw Figures 3 and 4: after the evolutionary dynamics has settled to an equilibrium (i.e., all agents are using the same communication code, except for single temporary mutants), the resulting homogeneous population is then left to interact for 100 contests and the average payoff is recorded. However, instead of exhibiting the payoff obtained in the 100 independent runs as in Figure 3, we exhibit in Figure 5 only the average payoff calculated over those runs. Hence to obtain each data point of this figure we need to generate a set of data similar to that used to draw Figure 3. We choose as the independent variable the ratio between the

payoffs for inferring a neighbor of the correct meaning and the correct meaning ($r$), which can be interpreted also as a selective pressure for evolving compositional codes. For the sake of comparison, Figure 5 also shows the average payoffs of perfect compositional and random Saussurean codes.

The results in Figure 5 indicate that for $r = 0$ the performance of the communication codes, regardless of whether random, compositional or evolved, are identical. Explicitly, in this case we find $G = 1 - \varepsilon$ for any one-to-one mapping. Since the search space is restricted to the space of permutation matrices, it is not a surprise that the payoffs of the Saussurean codes serve as lower bounds to those of the evolved codes. This trivial finding should not be confused with the unexpected result exhibited in Figure 3 that the payoffs of the Saussurean codes serve as upper bounds to the payoffs of the evolved codes when the search space is enlarged to cover all binary language matrices. The results in Figure 5 show clearly that, despite the fact that compositionality can greatly improve the communication payoff of the population (see upper straight line in that figure), the evolved codes fall short of taking full advantage of the structure of the meaning-signal space to cope with the noise in the communication. As a result, the evolved codes are far from the optimal, perfect compositional codes, although they fare better than the Saussurean codes. Figure 6 explains the reason for that: the evolutionary dynamics actually succeeded to produce partially compositional codes, reducing thus the deleterious effects of noise.

It is interesting that the payoffs of the Saussurean codes increase when the pressure for compositionality increases (see Figure 5 and equation (8)), although they remain largely non-compositional in average (see Figure 6). The key to the explanation of this result is found in Figure 4 where we can see that half of the samples of the random Saussurean codes exhibit a positive value of the compositionality, which is then associated to a payoff value greater than $1 - \varepsilon$ ($= 0.8$ in that case) while the representatives of the other half have a payoff of $1 - \varepsilon$ at worst. It is clear thus that the resulting average payoff must be an increasing function of $r$.

The reason why the evolutionary dynamics failed to produce perfect compositional codes, despite their obvious advantage to cope with noisy signals, is that once a non-optimal communication code has become fixed (or even almost fixed) in the population, mutants carrying better codes cannot invade. In fact, those mutants will most certainly do badly when communicating with the resident agents and, as a result, will quickly be removed from the population. As pointed out, this is essentially the Allee effect of population dynamics.
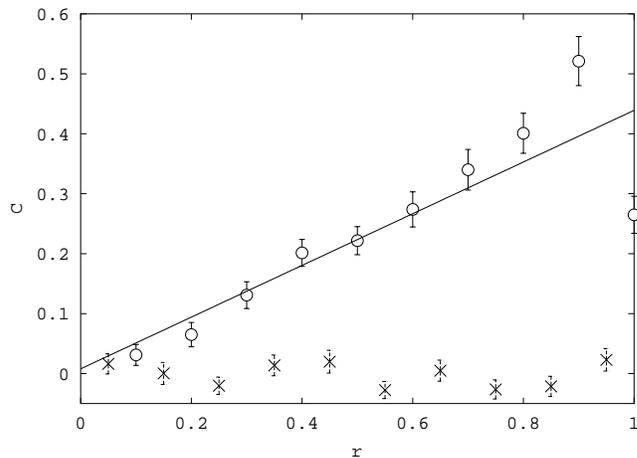


Figure 6: Average compositionality of the 100 evolved communication codes ($\circ$) whose payoffs are exhibited in Figure 5, as well as of the same number of Saussurean codes ($\times$). The compositionality of a perfect compositional code is $C = 1$ by definition. The linear fitting of the average compositionality of the evolved codes yields a slope of $\approx 0.43$.

The task faced by the evolutionary algorithm here is of an essentially different nature from that tackled in typical optimization problems in which the fitness of an agent is fixed a priori. In such case a fitter mutant always invades the resident population and thus guarantees that the optimum will eventually be found by the algorithm. To stress this phenomenon, Figure 7 illustrates the competition between a fraction $f$ of agents carrying (the same) perfect compositional code and a fraction $1 - f$ of agents carrying (the same) Saussurean code. This simulation is implemented using the elitist procedure described before, except that learning errors are not allowed, so that at any time an agent can carry only one of the two types of codes set initially. Alternatively, Figure 7 can be interpreted as the competition between two different strategies: the perfect compositional and the holistic strategies. We can easily estimate the minimum fraction $f_m$ of perfect compositional codes above which this strategy dominates the population. It is simply

$$\frac{f_m}{1 - f_m} = \frac{G_S}{G_c} \tag{9}$$

with $G_c$ and $G_S$ given by equations (7) and (8), respectively. For the parameters of Figure 8 this estimate yields $f_m \approx 0.46$, which, within statistical errors, is in very good agreement with the single run experiment described in the Figure. Repetition of this experiment using Moran's model rather than the elitist strategy leads to the same result, except that the fixation of the winner strategy takes much longer - about 100 times longer than the fixation times exhibited in Figure 7.
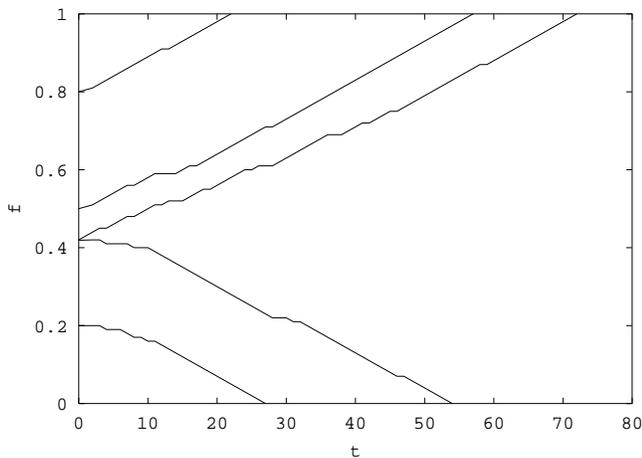
Figure 7: The evolution of the fraction $f$ of agents carrying a perfect compositional code in an experiment in which they compete against agents carrying a Saussurean code. The parameters are $\varepsilon = 0.5$, $r = 0.25$, $N = 100$ and $n = m = 10$. The initial population is set so that (from top to bottom) $f = 0.8, 0.5, 0.42, 0.419$ and $0.2$.

This simple analysis of the competition between suboptimal Saussurean codes and the optimal compositional codes lends support to our previous conclusion that compositional codes do not evolve within the usual language evolutionary game framework because the evolutionary dynamics is very likely to get trapped in the local maxima – the Saussurean codes.

## IV. INCREMENTAL MEANING ASSIMILATION

What we have been trying to do up to now is to evolve in a single shot a communication code that associates each one of the $n$ meanings (or objects) to one of the $m$ signals available in the repertoire of the agents. As pointed out, in the case that the meaning-signal mapping has a nontrivial underlying structure, this association is not completely arbitrary in the sense that in the presence of noise some codes (i.e., the perfect compositional codes) result in a much better communication accuracy than codes that implement an arbitrary one-to-one correspondence between meaning and signals (Saussurean codes). The results of the previous simulations lead us to conclude that it is very unlikely, if not impossible, that evolution through natural selection alone could take advantage of the structure of the meaning-signal space so as to produce the optimal, perfect compositional codes.

The outcome would be very different, however, if the task posed to the population were to reach a consensus on the signals to be assigned to the meanings in a sequential manner. In other words, let us consider the situation in which each agent has $m$ signals available (here we set $m = 10$) and the population needs to communicate about a single meaning, say $i = 1$. The search space is reduced then to the space of the $1 \times m$ permutation matrices. (We restrict the search space to that of permutation matrices, for simplicity.) Once the consensus is reached (i.e., the signal assigned to meaning $i = 1$ is fixed in the population), a new meaning is presented and the population is then challenged to find a consensus

signal for that meaning. The procedure is repeated until each one of the $n = m$ meanings are associated to a unique signal.

The order of presentation of meanings to the population plays a crucial role on the outcome of this strategy, which we term sequential meaning assimilation. In particular, success is guaranteed only if the novel meaning is chosen to be a neighbor of the previously presented meaning (e.g., $i = 2$ or $i = N$ in the case the previous assimilated meaning was $i = 1$). In this case, the question is whether the population will reach a consensus on a signal that is also a neighbor of the signal assigned to the previous meaning. Curve (a) of Fig. 8 shows that this scheme works neatly, and yields a fully compositional code provided that $\varepsilon \neq 0$ and $r \neq 0$. Note that the payoff of the sequential assimilation scheme (curve (a)) is below the average payoff a fully compositional code (dashed horizontal line) for $n < m$, although the codes produced by that scheme do take advantage of the topology of the meaning and signal spaces. This is so because the cyclic geometry of those spaces is not manifested until $n = m$. As a result, the agents get no reward if the noise corrupted signal is not associated with any of the previously assimilated meanings. For example, consider the situation in which two meanings were assimilated, say $i = 1, 2$ and the signals assigned to them were $j = 6, 7$, respectively. Clearly, there will be no reward if the corrupted signals become $5$ or $8$ (we recall that $m = 10$ in this experiment), whereas reward is always guaranteed for the fully formed compositional code. Of course, as seen in Fig. 8, this "surface" effect is attenuated as more meanings are assimilated. The fact that the final payoff of the single run displayed in curve (a) ends up being greater than the (theoretical) average payoff of the perfect compositional code is simply a statistical fluctuation. Curve (c) in Fig. 8 illustrates the failure of the sequential presentation scheme when the order of presentation of meanings is random. In fact, if the meanings are presented in an arbitrary order, say $i = 3$ after $i = 1$, then there is no selection pressure to prevent that the signal assigned to $i = 3$ be a neighbor of the signal associated to $i = 1$. Eventually, when the meaning $i = 2$ is presented this optimal signal will be unavailable to the agents, precluding thus the emergence of a compositional code. Finally, we note that the incremental learning scheme would work all the same if the repertoire of meanings were left fixed and the signals were presented one by one.
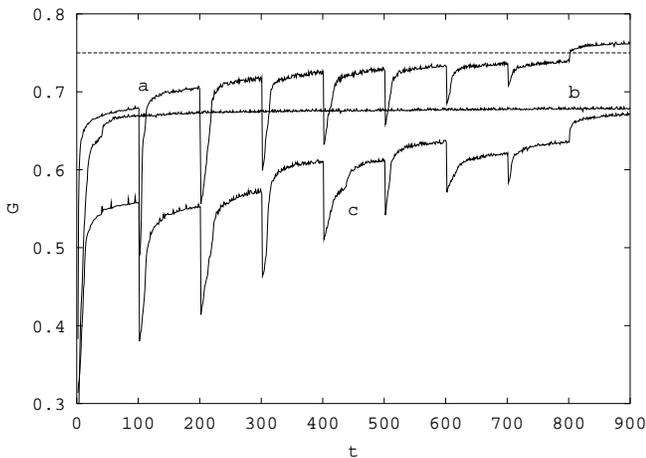
Figure 8: Average payoff of the population when the task is to produce consensus signals to $n$ meanings presented sequentially at the time intervals $\Delta t = 100$. In curve (a) the new meaning is a neighbor of the previous one, whereas in curve (c) the order of presentation of the meanings is random. The result for the usual batch algorithm, in which all meanings are presented simultaneously is shown in curve (b). The dashed horizontal line indicates the average performance of perfect compositional codes. The parameters are $\varepsilon = 0.5$, $r = 0.5$, $N = 100$ and $n = m = 10$.

The proposed solution to the evolution of compositional codes in an evolutionary language game framework could be questioned, because it relies on the assumption that the new meanings entering the population repertoire must be closely related to the already assimilated meanings. However, this seems to be the manner the perceptual systems work during categorization: new meanings are usually hierarchically related to the assimilated ones and this could be, for instance, the reason for Zipf's law of languages [40], [41]. In fact, as pointed out in [33], the hierarchical structure of language may be caused by our perception of reality, rather than the other way around. The case for a hierarchically organized world was made by Simon [42]: "On theoretical grounds we could expect complex systems to be hierarchies in a world in which complexity had to evolve from simplicity." In addition, the evidence that nouns are easily changed into verbs (e.g., ship – shipped, bottle – bottled) [43] illustrates the same type of continuity in the signal space as well.

In any event, our solution is in line with the traditional Darwinian explanation to the evolution of the so-called irreducibly complex systems. Although the evolutionary game setting failed to evolve perfect compositional codes when the task was to produce a meaning-signal mapping by assimilating all meanings simultaneously, that setting proved successful when the meanings were created gradually.

## V. CONCLUSION

Saussure's notion of language as a contract signed by members of a community to arbitrarily set the correspondence between words and meanings leads to unexpected obstacles to the evolution of efficient communication codes in the evolutionary language game framework. In fact, the fixation of a communication code in a population is a once-for-all

decision – it cannot be changed even if a small fraction of the population acquires a different, more efficient code (see Figure 7). The situation here is similar to the Nash equilibrium of game theory [44], the escape from which is only possible if all players change their strategies simultaneously. Since such concerted, global changes are not part of the rules of the language game, there seems to be no way for the population to escape from non-optimal communication codes.

In fact, languages evolve. A branch of linguistics named glottochronology (the chronology of languages) suggests the rule of thumb that languages replace about 20 percent of their basic vocabulary every one thousand years [45]. The abovementioned difficulty of changing the communication code is not in the replacement of old signals by new ones, but in the assignment of different meanings to old signals and vice-versa. Of course, this would not be an issue if the evolutionary language game could lead the population to the optimal code (a perfectly compositional code, in our case); our simulations have shown that it always gets stuck in one of the local maxima that plague the search space. To point out this difficulty was, in fact, the main goal of the present contribution.

Our view of compositionality as the evolutionary stage following the settlement of simpler, unstructured communication codes, and the search for a continuous path connecting these two stages, led us to the same type of difficulties researchers working on a similarly elusive problem - the origin of life - have been struggling with for more than three decades [39]. For example, although the coordinated work of distinct genes is germane to the emergence of cells, it is still not clear how such an assemblage could be formed and maintained starting from selfish genes (see [46] for a review). In that sense, by exposing the obstacles to explain compositionality from an evolutionary perspective, our work follows the same research vein that lead to the present understanding of pre-biotic evolution.

The solution we put forward to this conundrum is a conservative one – we cannot explain the emergence of the entire meaning-signal mapping that displays the required compositional property via natural selection, but it is likely that the mapping was formed gradually with the addition of one meaning at each time. This gradual procedure, that we term incremental meaning creation, leads indeed to fully compositional codes (see Figure 8). It would be interesting to verify whether alternative, less conservative solutions such as the spatial localization of the agents, less than perfect metrics in meaning space, or the structuring of the population by age could lead to the dissolution of the language contract and so open an evolutionary pathway to more efficient communication codes.

REFERENCES

[1] F. de Saussure, *Course in General Linguistics*. Translated by Wade Baskin. New York: McGraw-Hill Book Company, 1966.
[2] M. A. Nowak, N. L. Komarova, P. Niyogi, "Computational and evolutionary aspects of language," *Nature*, vol. 417, pp. 611-617, 2002.

[3]  J. R. Hurford, "Biological evolution of the Saussurean sign as a component of the language acquisition device," *Lingua*, vol. 77, pp. 187-222, 1989.

[4]  B. J. MacLennan, "Synthetic ethology: an approach to the study of communication," In: C. G. Langton, C. Taylor, J. Doyne Farmer, S. Rasmussen (Eds.), *Artificial Life II*, SFI Studies in the Sciences of Complexity, vol. X . Addison-Wesley, Redwood City, pp. 631-658, 1991.

[5]  K. Smith, S. Kirby, H. Brighton, "Iterated Learning: a framework for the emergence of language," *Artificial Life*, vol. 9, pp. 371-386, 2003.

[6]  H. Brighton, "Compositional Syntax from Cultural Transmission," *Artificial Life*, vol. 8, pp. 25-54, 2002.

[7]  H. Brighton, K. Smith, and S. Kirby, "Language as an evolutionary system," *Phys. Life Rev.*, vol. 2, pp.177-226, 2005.

[8]  M. A. Nowak, D. C. Krakauer, "The evolution of language," *Proc. Natl. Acad. Sci. USA*, vol. 96, pp. 8028-8033, 1999.

[9]  W. Zuidema, "Optimal communication in a noisy and heterogeneous environment," In: W. Banzhaf, T. Christaller, P. Dittrich, J. T. Kim, J. Ziegler (Eds.), *Advances in Artificial Life: Proceedings of the 7th European Conference on Artificial Life (ECAL)*, Lecture Notes in Artificial Intelligence , vol. 2801. Springer, Berlin, pp. 553-563, 2003.

[10]  W. C. Allee, *Animal Aggregations. A Study in General Sociology*. University of Chicago Press, Chicago, 1931.

[11]  F. Courchamp, T. Clutton-Brock, B. Grenfell, "Inverse density dependence and the Allee effect," *Trends Ecol. Evol.*, vol. 14, pp. 405-410, 1999.

[12]  D. Bickerton*, Language & Species*. University of Chicago Press, Chicago, 1990.

[13]  I. Ulbaek,,"The origin of language and cognition," In: J. R. Hurford, M. Studdert-Kennedy, C. Knight (Eds.), *Approaches to the Evolution of Language*. Cambridge University Press. Cambridge, UK, pp. 30-43, 1998.

[14]  L. Steels, L., 1996. "Perceptually grounded meaning creation," In: M. Tokoro (Ed.), ICMAS-96 *Proceedings Second International Conference on Multi-Agent Systems*. AAAI Press, Menlo Park, pp. 338-344, 1996.

[15]  J. F. Fontanari, and L. I. Perlovsky, "Meaning creation and Modeling Field Theory," In: C. Thompson, H. Hexmoor (Eds.), *Proceedings International Conference on Integration of Knowledge Intensive Multi-Agent Systems*. IEEE Press, Boston, pp. 405-410, 2005.

[16]  M. A. Nowak, J. B. Plotkin, and D. C. Krakauer, "The Evolutionary Language Game," *J. theor. Biol.*, vol. 200, pp. 147-162, 1999.

[17]  J. B. Plotkin, M. A. Nowak, "Language evolution and information theory," *J. theor. Biol.*, vol. 205, pp. 147-159, 2000.

[18]  J. F. Fontanari, and L. I. Perlovsky, "Evolution of communication in a community of simple-minded agents," In: C. Thompson, H. Hexmoor (Eds.), *Proceedings International Conference on Integration of Knowledge Intensive Multi-Agent Systems*. IEEE Press, Boston, pp. 285-290, 2005.

[19]  D. K. Lewis*, Convention: a Philosophical Study*. Harvard University Press, Cambridge, MA, 1969.

[20]  R. Dawkins, J. R. Krebs, "Animal signals: information or manipulation?" In: J. R. Krebs, N. B. Davies (Eds.), *Behavioural ecology: an evolutionary approach*. Blackwel Scientific Publications, Oxford, UK, pp. 282-309, 1978.

[21]  J.-L. Dessalles, "Altruism, status and the origin of relevance," In: J. R. Hurford, M. Studdert-Kennedy, C. Knight (Eds.), *Approaches to the Evolution of Language*. Cambridge University Press. Cambridge, UK, pp. 130-147, 1998.

[22]  R. L. Trivers, "The evolution of reciprocal altruism," *Quart. Rev. Biol.*, vol. 46, pp. 35-57, 1971.

[23]  W. T. Fitch, "Kin selection and mother tongues: A neglected component in language evolution," In: K. Oller, U. Griebel (Eds.), *EvolutionA of Communication Systems: A Comparative Approach*. MIT Press, Cambridge, MA, pp. 275-296, 2004.

[24]  G. M. Burghardt, "Defining Communication," In: J. W. Johnston Jr., J. D .G. Moulton, A. Turk (Eds.), *Communication by Chemical Signals*. Appleton-Century-Crofts, New York, pp. 5-18, 1970.

[25]  M. D. Hauser, *The Evolution of Communication*. MIT Press, Cambridge, 1996.

[26]  M. Oliphant, "The dilemma of Saussurean communication," *BioSystems*, vol. 37, pp. 31-38, 1996.

[27]  J. Maynard Smith, *Evolution and the Theory of Games*. Cambridge University Press, Cambridge, UK, 1982.

[28]  M. Mitchell, *An Introduction to Genetic Algorithms*. MIT Press, Cambridge, 1996.

[29]  M. A. Nowak, J. B. Plotkin, V. A. A. Jansen, "The Evolution of Syntactic Communication," *Nature*, vol. 404, pp. 495-498, 2000.

[30]  M. A. Nowak, N. L. Komarova, P. Niyogi, "The Evolution of Universal Grammar," *Science*, vol. 291, pp. 114-118, 2001.

[31]  S. Pinker, *The Language Instinct*. Penguin Press, London, 1994, pg. 152.

[32]  L. A. Petitto, "Language in the prelinguistic child," In: P. Bloom (Ed.), *Language acquisition: Core readings*. MIT/Bradford Press, Cambridge, 1994.

[33]  P. T. Schoenemann, "Syntax as an Emergent Characteristic of the Evolution of Semantic Complexity," *Minds and Machines*, vol. 9, pp. 309-346, 1999.

[34]  R. Goulden, P. Nation, J. Read, "How large can a receptive vocabulary be?" *Applied Linguistics*, vol. 11, pp. 341-363, 1990.

[35]  W. J. Ewens, *Mathematical Population Genetics*, 2nd edition. Springer, New York, 2004.

[36]  D. Fudenberg, J. Tirole, *Game Theory*. MIT Press, Cambridge, MA, 1991.

[37]  R. M. Seyfarth, D. L. Cheney, P. Marler, "Monkey responses to three different alarm calls: Evidence of predator classification and semantic classification," *Science*, vol. 210, pp. 801-803, 1980.

[38]  Lachmann, M., Bergstrom, C. T., "The disadvantage of combinatorial communication," *Proc. R. Soc. Lond. B*, vol. 271, pp. 2337-2343, 2004.

[39]  M. Eigen, "Self-organization of matter and the evolution of biological macro-molecules," *Naturwissenschaften*, vol. 58, pp. 465-523, 1971.

[40]  H. A. Simon, "On a class of Skew Distribution Functions," *Biometrika*, vol. 42, pp. 425-440, 1955.

[41]  L. B. Levitin, B. Schapiro, L. I. Perlovsky, "Zipf's Law Revisited: Evolutionary Model of Emergent Multiresolution Classification," *Proceedings of the Conference on Intelligent Systems and Semiotics '96*. Gaithersburg, MD, vol. 1, pp. 65-70, 1996.

[42]  H. A. Simon, *The Sciences of the Artificial*. MIT Press, Cambridge, MA, 1996.

[43]  P. J. Hopper, S. A. Thompson, "The discourse basis for lexical categories in universal grammar," *Language*, vol. 60, pp. 703-752, 1984.

[44]  J. Nash, "Equilibrium points in n-person games," *Proc. Natl. Acad. Sci. USA*, vol. 36, pp. 48-49, 1950.

[45]  C. Renfrew, *Archaeology and Language*. Pimlico, London, 1998.

[46]  J. Maynard Smith, E. Szathmáry, *The major transitions in evolution*. Freeman, Oxford, 1995.

# ALLEE EFFECT ON LANGUAGE EVOLUTION

JOSÉ F. FONTANARI

*Instituto de Física de São Carlos, Universidade de São Paulo, Caixa Postal 369,*
*São Carlos, SP 13560-970, Brazil*

LEONID I. PERLOVSKY

*Air Force Research  Laboratory, 80 Scott Rd., Hanscom Air Force Base, MA*
*01731, USA*

The case for the  study of the evolution of communication[a] within a multi-agent framework was  probably best made  by Ferdinand de Saussure in a famous statement made in his lectures at the University of Geneva (1906-1911) "language is not complete in any speaker; it exists only within a collectivity... only by virtue of a sort of contract signed by members of a community" (Saussure, 1966).  More than one decade ago, seminal computer simulations were carried out to demonstrate that natural selection (MacLennan, 1991) or, alternatively, learning  (Hurford, 1989) could lead to the emergence of ideal communication codes (i.e., one-to-one   correspondences between objects or meanings and signals) in a population of interacting agents. Typically, the behavior pattern of the agents was modeled by (probabilistic) finite state machines. The work by Hurford, in particular, set the basis of the celebrated Iterated Learning Model (ILM) for the cultural evolution of language (Smith et al, 2003).  In those studies, language is viewed as a mapping between meanings and signals. The abovementioned ideal codes that emerge from the agents interactions are examples of non-compositional or holistic communication, in which a signal stands for the meaning as whole. In contrast, a compositional language is a mapping that preserves neighborhood relationships – similar signals are mapped into similar meanings. The emergence of compositional languages in the ILM framework beginning from holistic ones in the presence of bottlenecks on cultural transmission was considered a major breakthrough in the computational language evolution field.  Our aim in this contribution is twofold. First, we  show that in practice, though contrasting at first sight, the cultural evolution approach in which the offspring *learn* their language from their

---

[a] Here we take the more conservative viewpoint that language evolved from animal communication rather than from animal cognition.

parents (or from other members of the community) differs very little from the genetic approach, in which the offspring *inherit* their communication ability from their parents. For instance, errors in the learning stage or the inventiveness associated to bottleneck transmission have the same effect of mutations in the genetic approach. Second, we show, through extensive simulations of language evolutionary games, that once an ideal communication code, say a holistic one, is established in the population, i.e., all individuals use the same code, it is impossible for a mutant to invade, even if the mutant uses a better code, say, a compositional one. This is essentially the Allee effect (Allee, 1931) of population dynamics which, for instance, prevents a population of asexual individuals of being invaded by a sexual mutant. The ILM circumvents this difficulty by assuming that the population is composed of two individuals only, the teacher and the pupil, and that the latter always replaces the former. However, according to Saussure (see quotation above), this is not an acceptable framework for language. The solution of the conundrum - how a compositional code can evolve in a population of agents that communicate through a holistic code - may give a clue on the interplay between cultural and genetic mechanisms in the evolution of language as well as support the viewpoint that language can in principle emerge from animal communication.

## References

Allee, W. C. (1931). *Animal Aggregations. A Study in General Sociology*. Chicago: University of Chicago Press.

de Saussure, F. (1966). *Course in General Linguistics*. Translated by Wade Baskin. New York: McGraw-Hill Book Company.

Hurford, J. R. (1989). Biological evolution of the Saussurean sign as a component of the language acquisition device. *Lingua*, *77*, 187-222.

MacLennan, B. J. (1991). Synthetic ethology: an approach to the study of communication. In Artificial Life II, SFI Studies in the Sciences of Complexity, vol. X (pp. 631-658). Redwood City: Addison-Wesley.

Smith, K., Kirby, S., Brighton, H. (2003). Iterated Learning: a framework for the emergence of language. *Artificial Life*, 9, 371-386.

# HOW LANGUAGE CAN GUIDE INTELLIGENCE

LEONID I. PERLOVSKY

*Air Force Research Laboratory, 80 Scott Rd., Hanscom Air Force Base, MA
01731, USA*


JOSÉ F. FONTANARI

*Instituto de Física de São Carlos, Universidade de São Paulo, Caixa Postal 369,
São Carlos, SP 13560-970, Brazil*

Today the favored explanation for the evolution of language seems to lie in the field of social intelligence. According to this view, language developed as a social glue: the primary selective pressure being the binding together of the early hominids in large groups, with gossip substituting costly grooming as the main mechanism of social interaction and cohesion (Dunbar, 1998). Nevertheless, advancing the argument that, taking language away, human social life may not be more complex than those of chimpanzees and bonobos, Calvin & Bickerton (2000) have championed the viewpoint that the selective pressures for language must have come from the brute exigencies of survival, e.g., hunting, food gathering and predator detection, rather than from human social life. Here we build on this proposal by considering these elementary survival needs as problems to be solved by the (artificial, in our case) organisms and ask how and whether communication can improve the performance of the individual organisms to solve a specific problem. This approach is in line with the seditious view of language as the cause of our species becoming more intelligent rather than that language being an inevitable consequence of greater intelligence.

The specific task we consider in this contribution is the differentiation problem, i.e., how organisms develop a more detailed knowledge of their surroundings. In particular, we address the problem of the "true" number of objects in the world, which is described as follows. We assume that the world contains a certain number of objects, e.g., points on a single axis or sets of points drawn from a Gaussian distribution, and that the organisms are endowed with a categorization system inspired in the modeling field theory (MFT) approach (Perlovsky, 2001) that, in principle, enables them to distinguish, through the creation of internal representations or concepts, those objects. At the beginning each organism starts with a single concept-model – a modeling

neuronal field chosen randomly - which then becomes associated to a specific object or group of objects. The organisms then exchange information – the values of their models or, alternatively, signs (words) associated to those models – which prompt them to create new concept models and finally to identify unambiguously all objects. We discuss the trade-off between the number of objects and the number of organisms needed to achieve perfect categorization. In doing so we demonstrate that categorization is better (in the sense that all objects are identified) and faster when communication is allowed.

This formulation allows us to go beyond the simplistic view of language as a mapping between objects in the real world and words (or, alternatively, between conceptual representations – meanings - and words) that underlies most of the simulation models on the evolution of language. In fact, since de Saussure it is known that there are at least two mapping operations between the real world and language: first our sense perceptions are mapped onto a conceptual representation, and then this conceptual representation is mapped onto a linguistic representation (Bickerton, 1990). The importance of the incorporation of this second hierarchy level in models for language evolution is the fact that linguistic representations can help creating conceptual categories , which may aid in coping with the external world. Another approach, that also shows the benefit of language to solve tasks that require the coordinated action of distinct agents, is the Predator-Prey Pursuit Problem (see, e.g., Jim & Giles, 2000). However, rather than provide additional support to this hardly surprising finding, our aim here is to verify the emergence of improved structure in combined categorization and communication abilities when the more realistic two-steps mapping between objects and words is implemented through the MFT formalism.

## References

Dunbar, R. (1998). *Grooming, Gossip, and the Evolution of Language*. Cambridge: Harvard University Press.

Bickerton, D. (1990). *Language & Species*. Chicago: University of Chicago Press.

Calvin, W. H., & Bickerton, D. (2000). *Lingua ex Machina*. Cambridge: MIT Press.

Jim, K.-C., & Giles, C. L. (2000). Talking Helps: Evolving Communication Agents for the Predator-Prey Pursuit Problem. *Artificial Life, 6,* 237-254.

Perlovsky, L. I. (2001). *Neural Networks and Intellect: Using Model-Based Concepts*. Oxford: Oxford University Press.

# Categorization and symbol grounding in a complex environment

José F. Fontanari, and Leonid I. Perlovsky, *Member, IEEE*

*Abstract*—**In order that communication can take place there must be something to be communicated. This basic stage of language evolution is the symbol grounding problem which addresses the issue of how physical signs acquire meaning. It is the symbols (e.g., words) associated to those meanings that are communicated by language. Here we show how the combination of the Modeling Field Theory and the Akaike Information Criterion can be used to find the true number of objects in an environment. We demonstrate creation of suitable representations and meanings for those objects and discuss the possible role of language in improving these representations.**

## I. INTRODUCTION

COMMUNICATION is not what language is, but what language does! This claim underlies the powerful view, championed by the linguist Derek Bickerton [1], [2] that language is primarily a representational system, established well before our remote ancestors have uttered the first recognizable word. Although in such a latent form language would be essentially unusable – only through communication language could have progressed from latency to its present status - there is no impediment, in principle, that individuals endowed with such a representational system could have invented purely mental labels for the categories they created, thus benefiting from (symbolic) thought even without making it public through communication.

It is on this pre-communication stage of language or protolanguage that we will focus in this contribution. In addition to being of interest on its own in the language evolution context as pointed above, the cognitive task of giving labels to categories is directly related to the symbol grounding problem [3]-[6] that addresses the question of how physical signs (e.g., gestures and sounds) can be given meaning. Although the symbol grounding problem may be better placed in the realm of cognitive rather than linguistic abilities, it represents a major challenge that must be addressed by any theory that purports to explain the origins of language [5]. In fact, language is not an isolated capability of the individual and cannot be fully comprehended if one ignores its intrinsic relationships with

J. F. Fontanari is with Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, SP Brazil (phone: +55-16-33739849; fax: +55-16-33739877; e-mail: fontanari@ifsc.usp.br).

L. I. Perlovsky is with Air Force Research Laboratory, Hanscom Air Force Base, MA 01731 USA (e-mail: Leonid.Perlovsky@hanscom.af.mil).

the cognitive and social abilities [6].

In this contribution meaning is viewed as a categorization of reality which is relevant from the perspective of the individual. Meaning creation is thus synonymous to category creation, i.e., the ability to distinguish, through the creation of internal representations or concepts, the objects, as well as the other individuals, that make up the individual's *Umwelt* (ethologist's jargon for the environment in which an individual is embodied and embedded).

A minimal model to study the meaning creation and the symbol grounding problems was proposed by Luc Steels [4], [5] and a simplified version of it is described as follows. (We refer the reader to [7] for a recent, lucid debate on the assumptions of Steels' approach.) An individual inhabits a simple world made up of $N$ objects or situations, each of which is described by a single feature value modeled by a real variable $O_i \in (0,1), i = 1, \cdots, N$ drawn randomly from some probability distribution. We note that in the original proposal [4], [5] each object is characterized by a set of features and each individual has a set of sensory channels designed to detect each feature (there is a one-to-one mapping between channels and features). Here we assume that there is only one feature per object and that the individuals possess a single sensory channel sensitive to that feature value. These features are, of course, abstract and have no particular meaning in the model, though it may be helpful to think of them as perceptual features such as color or smell. The question is whether such individual is able to form *autonomously* a repertoire of categories to succeed in discrimination and to adapt that repertoire when new objects are considered.

In his seminal works, Steels has tackled this issue using the so-called discrimination games, which may be viewed as a generalization of Wittgenstein's language games [8] to the non-linguistic domain. More specifically, a binary discrimination tree is introduced whose leaves (i.e., external nodes) are sensitive to certain ranges of the features values that describe the objects. If the existent leaves are not sufficient to distinguish between two objects, then the discrimination game fails and one randomly chosen leaf is split into two new nodes in order to increase the discrimination capability of the tree. Although this random refinement procedure eventually produces a discrimination tree capable of distinguishing between all $N$ objects, the finding that the number of leaves of such a successful tree increases exponentially with $N$ reduces considerably the applicability of this scheme to real-world situations [9]. In this simple single-channel scenario, the meaning of (or the symbol associated to) a given object is the unique leaf

sensitive to that object feature value.

In this contribution we address the symbol grounding problem as posed above using a novel adaptive approach to concept formation, Modeling Field Theory (MFT) [10]. In particular, extending our previous work on this theme [9], here we address a more difficult question than the mere categorization of the different objects in a number of classes, namely, how does an individual decide how many concepts are needed to account for the stimuli coming from the external world? In other words, how many objects are in the world? A biological organism evolves various complex mechanisms, related to instinctual and emotional evaluations, to make such a decision, i.e., to distinguish between the objects and the meaningless background that compose its world. An adaptation of a quote by Ferdinand de Saussure may be appropriate to describe this situation – without labels the world is a vague, uncharted nebula. (The original quotation is "Without language, thought is a vague, uncharted nebula. There are no pre-existing ideas, and nothing is distinct before the appearance of language" [11]. It is amazing how well this excerpt fits the notion that language is primarily a representational system.) But too many labels are equivalent to have no labels at all. In fact, mathematical approaches to determine the true number of objects are nontrivial because any data can be better fitted with more models (i.e., concepts). Here we will show how the problem of determining the "true" number of objects in the world can be approached by combining the Modeling Field Theory framework with the Akaike Information Criterion [12], [13] to penalize solutions that use too many models.

## II. THE MODELING FIELD THEORY FRAMEWORK

The basic idea behind Modeling Field Theory is the association between lower-level signals (e.g., inputs, bottom-up signals) and higher-level concept-models (internal representations, top-down signals) avoiding the combinatorial complexity inherent to such a task. This is achieved by using measures of similarity between concept-models and input signals together with a new type of logic, so-called dynamic logic. We refer the reader to [10] for a complete presentation of MFT; here we particularize the general framework to the problem of categorizing $N$ objects, each of which is characterized by a real number $O_i \in (0,1)$ - the input signals - as described in the previous section. Let us start with $M$ concept-models, or neuronal fields, described by real-valued variables $S_k, k = 1, \cdots, M$ that should represent the objects $O_i, i = 1, \cdots, N$. We use the following partial similarity measure [10] between object $i$ and concept $k$

$$l(i \mid k) = \left(2\pi\sigma_k^2\right)^{-1/2} \exp\left[-\left(O_i - S_k\right)^2 / 2\sigma_k^2\right] \qquad (1)$$

where, at this stage, the fuzziness $\sigma_k$ is a parameter given *a priori*. The goal is to find an assignment between models and objects such that the global similarity

$$L = \frac{1}{M} \sum_i \log \sum_k l(i \mid k) \qquad (2)$$

is maximized. For our purposes, namely, to compare the values of $L$ obtained using distinct number of model-fields, it is germane that we re-normalize the global similarity by the number of fields, as done in (2), in order to make it an intensive quantity with respect to $M$.

The maximization of $L$ can be achieved using the MFT mechanism of concept formation which is obtained through the direct maximization of (2) with respect to $S_k$. The aim here is to derive a dynamical equation for the modeling fields $S_k$ such that $dL/dt \geq 0$ for all time $t$. This condition can easily be met by choosing $dS_k/dt = \partial L/\partial S_k$ since then

$$dL/dt = \sum_k \left(\partial L/\partial S_k\right)\left(dS_k/dt\right) = \sum_k \left(\partial L/\partial S_k\right)^2 \geq 0 \qquad (3)$$

as required. The calculation of $\partial L/\partial S_k$ is straightforward

$$\frac{\partial L}{\partial S_k} = \frac{1}{M} \sum_i \frac{1}{\sum_{k'} l(i \mid k')} \frac{\partial l(i \mid k)}{\partial S_k} \qquad (4)$$

and leads to the following dynamics for the modeling fields

$$dS_k/dt' = \sum_i f(k \mid i)\left[\partial \log l(i \mid k)/\partial S_k\right], \qquad (5)$$

where we have used the identity $\partial y/\partial x = y \partial \log y/\partial x$ and re-scaled the time $t' = t/M$. (Henceforth we will drop the prime mark in $t'$ for simplicity of notation.) The fuzzy association variables $f(k \mid i)$ are defined by

$$f(k \mid i) = l(i \mid k) \Big/ \sum_{k'} l(i \mid k'), \qquad (6)$$

and give a measure of the correspondence between object $i$ and concept $k$ relative to all other concepts $k'$. It can be shown that this dynamics always converges to a (possibly local) maximum of the similarity $L$ [10]. By properly adjusting the fuzziness $\sigma_k$ the global maximum can be attained (see, however, Section III). A salient feature of dynamic logic is a match between parameter uncertainty and fuzziness of similarity. In what follows we decrease the fuzziness during the time evolution of the modeling fields according to the following prescription

$$\sigma_k^2(t) = \sigma_{k1}^2 \exp(-\alpha t) + \sigma_{k0}^2 \qquad (7)$$

with $\alpha = 5 \times 10^{-4}$, $\sigma_{k1} = 1$ and $\sigma_{k0} = 0.03$ for $k = 1, \ldots, M$ so the variance of the Gaussian similarity measure (1) becomes model-independent. Unless stated otherwise, these are the parameters we will use in the forthcoming analysis. In [9] we have shown that this setting allows perfect categorization, in a sense that the values of the modeling fields match those of the objects, provided that the number of modeling fields $M$ is equal or greater than the number of

objects $N$. As a guideline for setting the parameter values in (7) we note that $\sigma_{k1}$ must be chosen large enough such that, at the beginning, all objects are described by all fields, whereas the baseline resolution $\sigma_{k0}$ must be small enough such that, at the end, a given field will describe a single object. However, $\sigma_{k0}$ should not be set to a too small value to avoid numerical instabilities in the calculation of the partial similarities (1).

A word is in order about the connection between the MFT and neural networks. A MFT neural architecture was described in [10], which combines architecture with models of objects. Essentially, input neurons or bottom-up signals encode the object feature values $O_i$, and top-down or priming signal-fields to these neurons are generated by the modeling fields $S_k$. Interaction between bottom-up and top-down signals is determined by the neural weights $f(k\,|\,i)$ that associate signals and models. As described before, these weights are functions of the model parameters $S_k$, which in turn are dynamically adjusted so as to maximize the overall similarity between objects and models. This formulation sets MFT apart from many other neural networks. There is, on the other hand, a certain formal similarity between the MFT approach and the Hopfield-Tank neural network approach to tackle optimization problems [14]. This becomes apparent when one recognizes that the nature of perceptual problems dealt with here is similar to that of other optimization problems. In fact, in both systems it is the time evolution of analog neurons that drives the neural configuration to a maximum of the cost function [the global similarity (1) in our case]. In addition, the quality of the solutions found by the neural network is greatly improved by annealing the analog gain parameter [15], in a similar manner as the slow decrease of the fuzziness according to (7) leads ultimately to perfect categorization.

The MFT framework alone, however, does not account for the need to decide how many different models (i.e., modeling fields) the organism really needs. Therefore it is necessary to balance maximization of similarity as given in (2), against the number of parameters in the model. A theoretically consistent way to achieve this balance is to use Akaike Information Criterion, AIC for short, which is an asymptotic correction to the similarity function related to the bias due to the number of parameters, namely [12],

$$AIC = L - M_{par} \qquad (8)$$

where $M_{par}$ is the number of adjustable parameters of the models, and $L$ is the likelihood function given by (2). Since here the models are defined by a single parameter ($S_k$) we have $M_{par} = M$. Note that the fuzziness $\sigma_k$ is not considered a parameter of the model – it is simply a parameter that appears in the functional form of the partial similarities measures, regardless of the choice of the model. The basic idea behind the AIC methodology is to analyse the complexity of a model, as given by the number of adjustable parameters, together with the goodness of its fit to

the input data, and to produce a measure that balances between these two quality factors. Although a model with many parameters may provide a very good fit to the data, it will have little predictive value. This balanced approach thus inhibits overfitting. The preferred model is that with the highest AIC value. The general applicability and simplicity of the AIC for model selection prompted its use in a variety of areas such as hydrology, geophysics, engineering, econometrics, medicine and bioinformatics (see [13] for a recent review). In the following we apply this framework to identify the number of objects in a very simple case in which each object is represented by a single point in the real axis, and in a more complex situation in which the objects are clouds of points drawn from a Gaussian distribution.
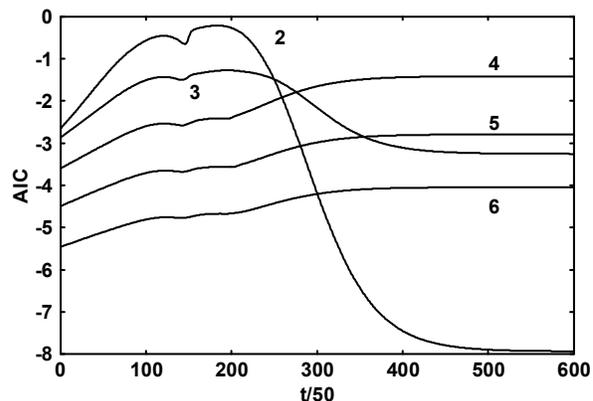


Fig. 1. Illustration of the use of Akaike Information Criterion (AIC) measure in conjunction with the MFT scheme with $M$ = 2, 3, 4, 5 and 6 modeling fields to determine the number of objects in the environment. Here the true number is $N$ = 4, which corresponds to the maximum of the AIC for large $t$.

### A. Simple objects

To better appreciate the effectiveness of the AIC to single out the true number of objects in the environment we consider a very simple situation in which there are $N$ = 4 objects: $O_1 = 0.2$, $O_2 = 0.4$, $O_3 = 0.6$ and $O_4 = 0.8$. The modeling field dynamic equations (5) – (7) are then solved numerically with Euler's method using the step-size $h = 10^{-5}$ for several choices of $M$ and the resulting value of the AIC, as given by (8), is plotted against time $t$. The initial values of the modeling fields $S_k(t = 0)$ are chosen randomly in the range $(0,1)$. The results shown in Fig. 1 illustrate how tricky the determination of the true value of $N$ can be. In fact, for short times, the choice of fewer models than the true number yields the maximum value of AIC, but as the dynamics progresses the insufficiency of models becomes readily noticeable and, as expected, in the asymptotic regime $t \to \infty$ the maximum of AIC corresponds to the situation $M = N$. Interestingly, the observed decrease of AIC in the under-represented case $M < N$ yields a clear indication that something is going wrong, serving thus as a sign for increasing the number of models. On the other hand, by following the time evolution in the over-represented case

$M > N$ , say $M = 6$, we find no clue of the use of an excessive number of models, unless we explicitly compare AIC values for different numbers of models.
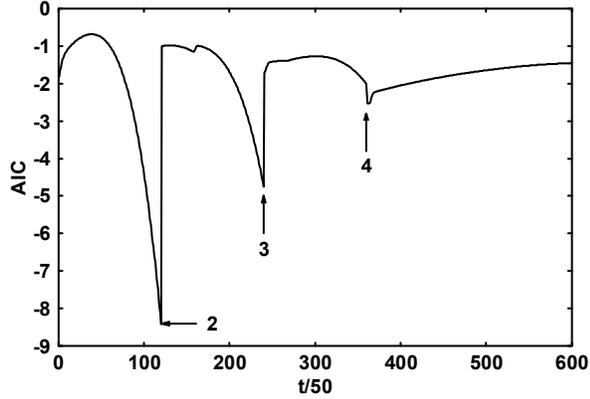


Fig. 2. Results of the adaptive scheme to find the true number of objects for the same problem of Fig. 1. Starting with a single model ($M$=1) the evolution of AIC measure is followed until a decrease is detected (this check is done at time intervals of $\Delta t$=3000) then a new model is created. The arrows indicate the moments when the second, third and fourth models are created.
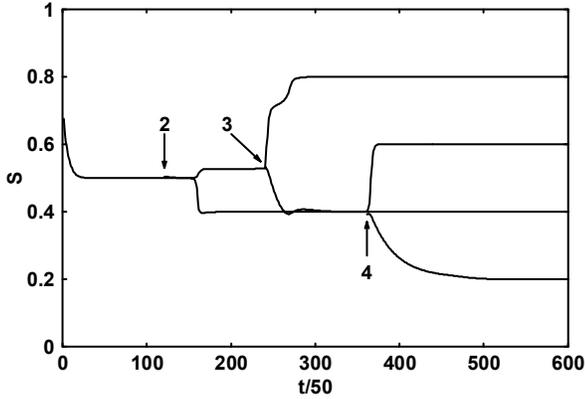


Fig. 3. Time evolution of the modeling fields using the adaptive scheme to create new fields on the fly based on the behavior pattern of the AIC. These data correspond to the same experiment depicted in the previous figure. To identify the fields we have only to note that a novel field is created as a perturbation of the previously created one. The final assignment is $S_1 = O_2$, $S_2 = O_4$, $S_3 = O_3$, and $S_4 = O_1$.

Taking advantage of the distinctive behavior pattern of the dependence of AIC on $t$ in the under-represented case, we envisage a simple strategy to adjust the value of $M$ on the fly: starting with a single model $S_1$, we create a new model whenever AIC decreases. The value of the new modeling field created at $t = t_c$, say $S_2(t_c)$, is then given by a perturbation of one of the previous fields, e.g., $S_2(t_c) = S_1(t_c) + 0.01\varepsilon$, where $\varepsilon$ is a random number drawn uniformly in the interval (-1,1). In addition, the fuzziness of the new model obeys the re-scaled equation (7), $\sigma_2^2(t) = \sigma_{21}^2 \exp[-\alpha(t - t_c)] + \sigma_{20}^2$. The trouble with this procedure is that by adding a new model that, in principle,

has a small similarity with all objects, we simultaneously decrease $L$ and increase $M_{par}$ in (8), which results in a further decrease of AIC. To circumvent this difficulty we must allow some time, i.e., a time interval $\Delta t = 3000$ , for the new field to adapt to the objects and only then to check for a decrease of AIC. The result of applying this strategy to the same categorization problem addressed in Fig. 1 is depicted in Fig. 2 and the details of the time evolution of the modeling fields are presented in Fig. 3.

At this stage it is appropriate to stress a certain similarity between the autonomous procedure described above to identify the true number of objects in the world and the more abstract Modeling Field Theory view of the mind [10] (see also [16]). According to that viewpoint, instincts, concepts and emotions are among the fundamental mechanisms of the mind, which has evolved to guarantee a better satisfaction of the basic instincts needed to survival. Instincts are like internal sensors that prompt the organism to take some action when the organism is at risk. In the present context, we might say that there is an instinct to increase the quantity AIC, defined in (8) - in other words, an instinct for knowledge [16]. In addition, the role of emotions within the mind is the evaluation of the concepts for the purpose of instinct satisfaction. Hence, the evaluation of the AIC and the detection of its unwanted decreasing behavior are done by emotional signals. Finally, conceptual-emotional understanding of the world leads to an action which in our case is the creation of novel models that aim, ultimately, to promote the increase of AIC.
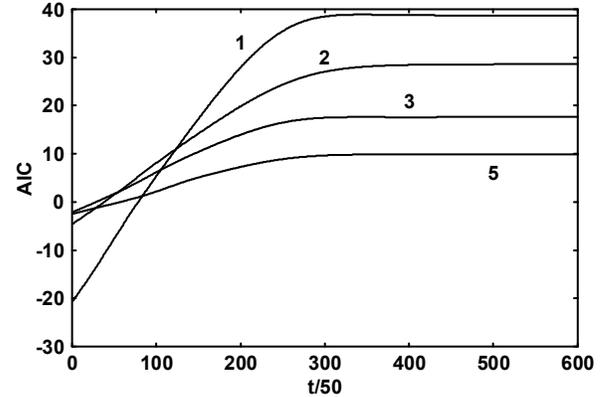


Fig. 4. Akaike Information Criterion (AIC) measure for $M$ = 1, 2, 3, and 5 modeling fields in the case that the world is composed of 20 points generated by the same Gaussian distribution. The true number of objects is $N = 1$, which corresponds to the maximum of the AIC for large $t$.

### B.    Complex Objects

Up to now we have considered the objects as points on a single axis. Here we assume that an object is a set of points drawn from a Gaussian distribution with mean $m$ and variance $v^2$ . The problem is to verify what conditions need to be satisfied in order that the MFT system recognizes the whole object and not the individual points that compose it. Of course, we expect that the final categorization ability of the system will depend strongly on the balance between the

baseline resolution of the modeling fields $\sigma_{k0}^2$, the variance $v^2$ and the distance between the means of the distributions associated to each object.

We begin with the simplest case in which there is a single object composed of 20 points generated according to a Gaussian distribution of mean $m = 0.5$ and standard deviation $v = 0.03$. Note that this is the same standard deviation associated to the baseline fuzziness $\sigma_{k0}^2 = 0.03^2$ of all models. The result shown in Fig. 4 indicates that for large $t$ the system is capable of identifying all points as parts of a single object. A similar, successful performance is obtained in a slightly more difficult problem (see Fig. 5) in which 40 points are drawn from two Gaussian distributions of means $m_1 = 0.3$ and $m_2 = 0.6$, and standard deviations $v_1 = v_2 = 0.03$.
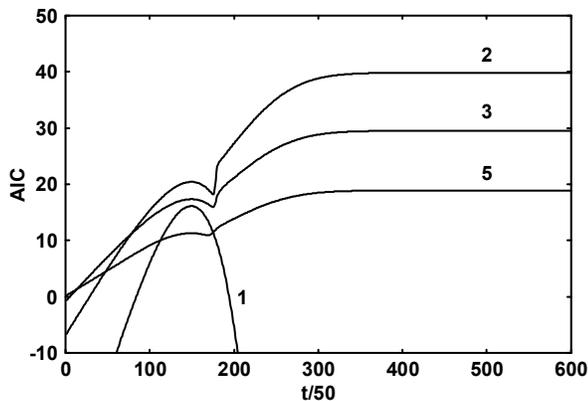


Fig. 5. Akaike Information Criterion (AIC) measure for $M = 1, 2, 3,$ and 5 modeling fields in the case that the world is composed of 40 points generated by two Gaussian distribution (20 points for each object). The true number of objects is $N = 2$, which corresponds to the maximum of the AIC for large $t$.
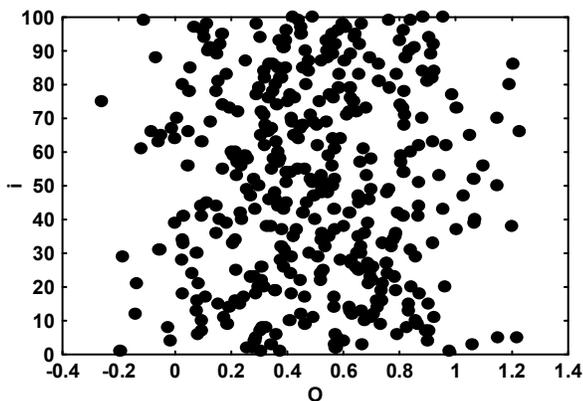


Fig. 6. The environment is composed of four objects each of which represented by 100 points drawn from Gaussian distributions of means 0.2, 0.4, 0.6, and 0.8, and standard deviations $v = 0.2$. All 400 points are plotted in the figure. For ease of visualization, the points are shown displaced vertically with four points per row. The original data is recovered by projecting all points into a single row. Would the reader be able to tell how many objects are displayed in the figure? This is the task set to our discrimination system.

Finally, we consider a more challenging situation that involves the discrimination of four overlapping objects, each of which represented by 100 points drawn from Gaussian distributions of means 0.2, 0.4, 0.6, and 0.8, and standard deviations equal to 0.2, as shown in Fig. 6.
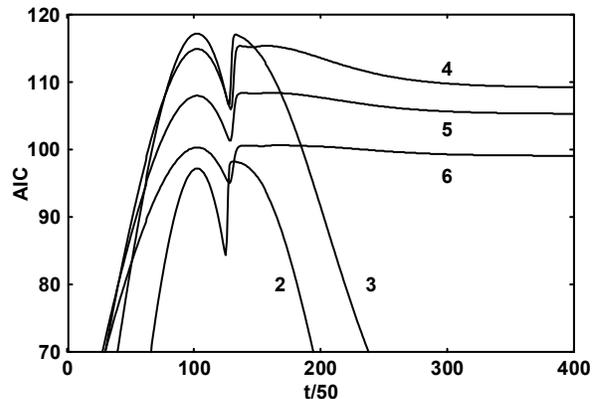


Fig. 7. Akaike Information Criterion (AIC) measure for $M = 2, 3, 4, 5$ and 6 modeling fields for the data displayed in Fig. 6. The true number of objects is $N = 4$, which corresponds to the maximum of the AIC for large $t$. The data for $M=1$ lay completely outside the scale of the figure since AIC reaches its maximum value $\approx -80$ at $t/50 = 70$.

The outcome of the application of our discrimination system to the data of Fig. 6 is presented in Fig. 7. Here, in order to guarantee the numerical stability of the differential equations we have set the baseline standard deviation to $\sigma_{k0} = 0.1$ for all models $k = 1, \ldots, M$. Surprisingly, maximization of the AIC measure for large $t$ yields the correct answer $M = 4$. However, the time dependence of this measure is very different from that observed in the simpler problems analyzed in Figs. 1, 4, and 5. In particular, there is a transient stage when the AIC measure increases until it reaches a maximum and then decreases towards a fixed value. This odd behavior pattern precludes the use of the automated scheme for generating new models we used to draw Figs. 2 and 3. It is instructive to follow the time evolution of the modeling fields in this more complex situation. This is shown in Figs. 8 and 9 for $M = 4$ and $M = 6$, respectively. From these figures we can see that the abrupt increase of the AIC measure that interrupts the smoothly decaying stage is associated to the simultaneous splitting of the modeling fields. The case of Fig. 9 is particularly interesting because it shows an additional merging of two models, which split again later.
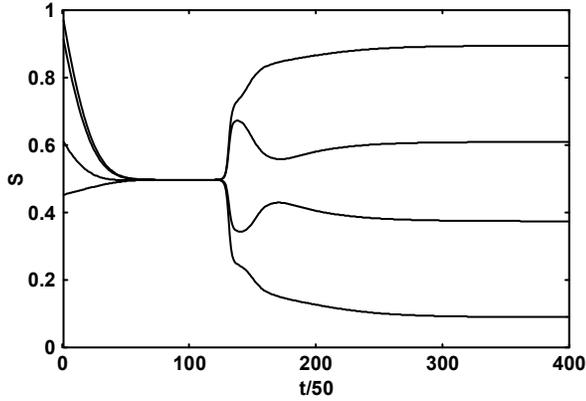
Fig. 8. Time evolution of the modeling fields for *M*=4. These data correspond to the same experiment depicted in the previous figure. Note that the asymptotic values of the fields do not match the means of Gaussians used to generate the points associated to the four objects.
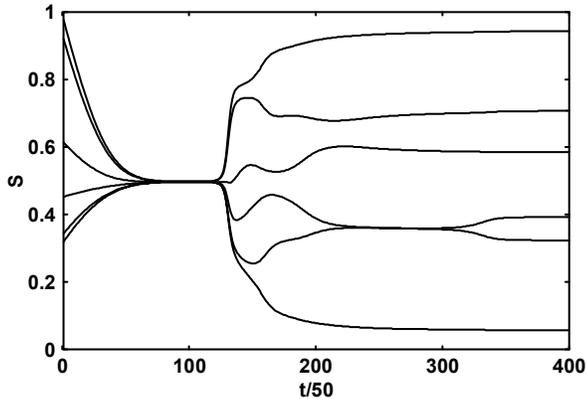


Fig. 10. Akaike Information Criterion (AIC) measure for *M* = 1, 2, 4, 6, 8 and 12 modeling fields for *N*=60 points distributed uniformly in (0,1). The asymptotic value of the AIC measure increases with the number of models *M* in stark contrast to the behavior pattern observed in structured environments.



Fig. 9. Time evolution of the modeling fields for *M*=6. These data correspond to the same experiment depicted in Figs. 6 and 7. The use of more (distinct) concepts results in a decrease of the AIC measure as compared to the correct guessing.

The experiment described in Figs. 6 to 9 yields a good indication of the potential of the framework that combines MFT with Akaike Information Criterion to identify objects in a complex environment. There is an important test, however, that must be done before we come to a definitive verdict on the usefulness of this discrimination system: what happens if the environment is completely unstructured, i.e., if the points are randomly scattered in the range $(0,1)$? The correct response in this case is to identify each point as a distinct object and this is exactly the tendency of the data depicted in Fig. 10 for $N = 60$ points distributed uniformly in the unit interval. It is difficult to increase much further the number of models *M* because we have to guarantee that the asymptotic values of the modeling fields are all distinct. Actually, to avoid the irreversible fusion of the modeling fields, in drawing Fig. 10 we have set $\sigma_{k0} = 0.01$ for all models. Nevertheless, the tendency of increasing the AIC measure with increasing *M* is very clear.
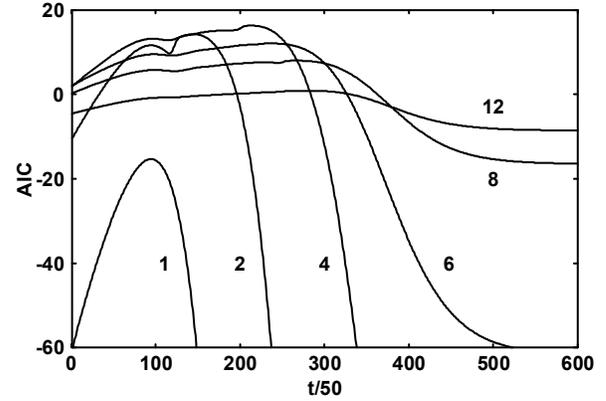
## III. GENERAL REMARKS

Looking at the time dependence of the AIC measure for *fixed M*, depicted in Figs. 1, 5, 7 and 10, immediately brings a question up: Shouldn't *L* as given in (2) (or, equivalently, AIC since *M* is kept fixed) be an increasing function of time? The answer is yes, provided that the fuzziness $\sigma_k, k = 1,…,M$ is kept fixed during the evolution of the fields $S_k$, which is not the procedure we are adopting here since (7) provides an explicit prescription for updating the fuzziness. Hence there is actually no reason to expect that *L* or the AIC measure will increase during the time evolution of the modeling fields. There is, however, a way to update the fuzziness so as to guarantee that *L* increases with increasing *t*: considering $\sigma_k$ as an adjustable parameter, similar to the modeling fields, we derive the equation

$$d\sigma_k/dt = \sum_i f(k \mid i)\left[\partial \log l(i \mid k)/\partial \sigma_k\right]. \qquad (9)$$

This equation solved simultaneously with (5) leads to increase of L until reaching the maximum. We have found that these equations tend to the uniform solution, i.e., $S_1 = S_2 = … = S_M$ and $\sigma_1 = \sigma_2 = … = \sigma_M$. Of course, such a solution is always a stable local maximum. In fact, inspection of Figs. 7 and 8 shows that, when using equation (7) instead of (9), an approximately homogeneous solution may yield a maximum of *L* at an intermediate point in the convergence process: the value of AIC at $t/50 = 100$ when the fields are all merged into a single field is greater than the AIC value of the more satisfactory asymptotic solution. This is so because the final baseline standard deviation $\sigma_0$ (in this case 0.1) is smaller than the true object standard deviation (in this case 0.2). This is an illustration of a general situation that a single field with a large standard deviation can account for most of the points in the environment – this is a stable, but unsatisfactory, solution for a difficult problem such as that posed in Fig. 9. In our setting the homogenous

solution breaks down because prescription (7) reduces continuously the fuzziness [use of (9) would allow the fuzziness to remain at a large value] so a single field can no longer account for all points in the environment. This is the reason why the categories initially merge into a single category and then split in the appropriate ones (see Figs. 8 and 9). This analysis indicates that if the exact variability (standard deviation) of objects is not known, more sophisticated approaches have to be explored.

Although for objects described by single points we have devised a scheme (or a sensor) for automatically creating new concepts (modeling fields) whenever the AIC measures decreases during a certain time interval, this scheme does not work in general as, for instance, in the problem illustrated in Fig. 7 since there the AIC measure decreases continuously as the satisfactory solution is approached. Of course, only if such a sensor is devised then one could say that the discrimination system is capable of inferring the true number of objects in the environment. Nevertheless our results indicate rather clearly that such a sensor can be based on the AIC measure.

## IV. CONCLUSION

This contribution follows the trend initiated in [9] by offering a series of didactic experiments to investigate the use of the Modeling Field Theory framework in solving complex aspects of categorization problem. In particular, here we show how the combination of that framework with the Akaike Information measure can be used to find the true number of objects in an environment, as well as to create a suitable representation for them. The difficulties with estimating the true number of different objects in the environment illustrated in this work are not new; they have been encountered in categorization research for years [17]. We expect that the information on the true number of the objects in the environment comes not merely from statistical properties of observed features, but from higher hierarchical levels in the organism, including communication among individuals. It is quite possible that multiple hierarchical levels of categorization require communication, and that reliable sophisticated categorization system appeared evolutionary together with language.

## REFERENCES

[1] D. Bickerton, *Language & Species*. University of Chicago Press, Chicago, 1990.

[2] D. Bickerton, *Language and Human Behavior*. University of Washington Press, Seattle, 1995.

[3] S. Harnard, "The symbol grounding problem," *Physica D*, vol. 42, pp. 335-346, 1990.

[4] L. Steels, "Perceptually grounded meaning creation," In: M. Tokoro (Ed.), ICMAS-96 *Proceedings Second International Conference on Multi-Agent Systems*. AAAI Press, Menlo Park, pp. 338-344, 1996.

[5] L. Steels, "Grounding symbols through evolutionary language games," In: A. Cangelosi, D. Parisi (Eds.), *Simulating the Evolution of Language*, London: Springer-Verlag, pp. 211-226, 2002.

[6] A. Cangelosi, A. Greco, and S. Harnard, "Symbol grounding and the symbolic theft hypothesis," In: A. Cangelosi, D. Parisi (Eds.), *Simulating the Evolution of Language*, London: Springer-Verlag, pp. 191-210, 2002.

[7] L. Steels and T. Belpaeme, "Coordinating perceptually grounded categories through language. A case study for colour," *Behavioral and Brain Sciences*, vol. 24, no. 8, pp. 469-529, 2005.

[8] L. Wittgenstein, *Philosophical Investigations*. Basil Blackwell, New York, 1968.

[9] J. F. Fontanari and L. I. Perlovsky, "Meaning creation and modeling field theory," In: C. Thompson, H. Hexmoor (Eds.), *Proceedings International Conference on Integration of Knowledge Intensive Multi-Agent Systems*. IEEE Press, Boston, pp. 405-410, 2005.

[10] L. I. Perlovsky, *Neural Networks and Intellect: Using Model-Based Concepts*. Oxford University Press, Oxford, 2001.

[11] F. de Saussure, *Course in General Linguistics*. Translated by Wade Baskin. McGraw-Hill Book Company, New York, 1966.

[12] H. Akaike, "A new look at the statistical model identification,"*IEEE Trans. Automat. Contr.*, vol. AC-19, pp. 716-723, 1974.

[13] K. P. Burnham and D. R. Anderson, *Model Selection and Multimodel Inference: A Practical-Theoretic Approach*. Springer, New York, 2002.

[14] J. J. Hopfield and D. W. Tank, "Neural computation of decisions in optimization problems," *Bio. Cybern.*, vol. 52, pp. 141-152, 1985.

[15] D. E. Vandenbout and T. K. Miller, "Improving the performance of the Hopfield-Tank neural network through normalization and annealing," *Bio. Cybern.*, vol. 62, pp. 129-139, 1989.

[16] L. I. Perlovsky, "Integrating language and cognition," *IEEE Connections*, vol. 2, no. 2, pp. 8-13, May 2004.

[17] V. Cherkassky and F. Mulier, *Learning from Data: Concepts, Theory, and Methods*. Wiley-Interscience, New York, 1998.

# Meaning creation and communication in a community of agents

José F. Fontanari, and Leonid I. Perlovsky, *Member, IEEE*

*Abstract*— The emergence of communication is studied in a scenario where agents endowed with distinct object-meaning mappings learn from scratch signal-meaning associations (i.e., communication codes) that allow them to identify the objects in their environment. Meanings are created through the Modeling Field Theory categorization mechanism, and learning is based on two variants of the obverter procedure, in which the agents may or may not receive feedback about the success of the communication episodes. We show that in the unsupervised learning scheme the agents fail to develop ideal communication codes, whereas success is guaranteed in the supervised scheme provided the size of the repertoire of signals is sufficiently large, though only a few signal are actually used in the code. Thus the mere ability to produce and observe different signals bears on the quality of the evolved communication codes.

## I. INTRODUCTION

LANGUAGE, according to [1], [2], is primarily a representational system, developed well before our remote ancestors have uttered the first recognizable word. Individuals endowed with such a representational system, it was hypothesized, could have invented purely mental labels for the categories they created, which according to the above references constituted symbolic thought. In such a dormant form, however, language would be essentially unusable – only through communication language could have evolved to become unarguably the most powerful representational system ever seen in nature. In fact, it is difficult to see what could be the benefits for an individual to mentally manipulate a few symbols, or in which way these symbols should be separate from other forms of mental representations, whereas the advantage of exchanging meaningful symbols with close relatives is plainly obvious.

In this contribution we offer a computational model that addresses both the meaning creation and the communication issues. Here meaning is viewed as a categorization of reality which is relevant from the perspective of the individual. Meaning creation is thus synonymous to category creation, i.e., the ability to distinguish the objects in the world through the creation of internal representations or private labels to those objects. How these labels are mapped into arbitrary

J. F. Fontanari is with Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, SP Brazil (phone: +55-16-33739849; fax: +55-16-33739877; e-mail: fontanari@ifsc.usp.br).

L. I. Perlovsky is with Air Force Research Laboratory, Hanscom Air Force Base, MA 01731 USA (e-mail: Leonid.Perlovsky@hanscom.af.mil).

signals that are then made available to other individuals (e.g., through sounds, gestures or chemical cues) and how these individuals infer their meanings constitute the issue of the origin of the communication.

Our model builds on the works of Steels [3]-[5] and Smith [6], [7] in that the architecture of the agents is composed of two parts, namely, a conceptualization module that embodies the categorization capability and a verbalization module that accounts for the transmission and reception of signals. In addition, similarly to those works, we do not allow the verbalization module to affect the conceptualization module, so the co-evolution of language and cognition is not addressed at this stage (see, e.g., [8]-[11] for contributions in this line, or [12] for the more extreme perspective that meaning emerges from communication).

There are, however, at least two significant differences between ours and the abovementioned approaches. First, the conceptualization module uses the Modeling Field Theory mechanism [13] to create the categories, rather than the Steels' discrimination trees [3]; and second, the inference procedure of the verbalization module uses a simplified variant of the *obverter* mechanism [14] that greatly facilitates the understanding of the method.

## II. MEANING CREATION

Here we consider the minimal model proposed by Steels to study meaning creation or symbol grounding [3], [4] which is described as follows (see [15] for a recent critical overview on this approach). An individual inhabits a simple world made up of $N$ objects, each of which is described by a single feature value modeled by a real variable $O_i \in (0,1), i = 1, \cdots, N$ drawn randomly from some probability distribution. These features are, of course, abstract and have no particular meaning in the model, though it may be helpful to think of them as perceptual features such as color or geometric form. The question is whether such individual is able to create *autonomously* a set of representations to succeed in discrimination and to adapt that set when new objects are considered.

To achieve that goal we use the Modeling Field Theory (MFT) framework [13] to produce the required associations between lower-level signals (e.g., inputs, bottom-up signals) and higher-level concept-models (internal representations, top-down signals). The MFT is based on measures of similarity between concept-models and input signals together with a new type of logic, so-called dynamic logic. We refer the reader to [13] for a complete presentation of

MFT; here we particularize the general framework to the problem of categorizing $N$ objects, each of which is characterized by a real number $O_i \in (0,1)$ - the input signals.

We introduce $M$ concept-models, or neuronal fields, described by real-valued variables $S_k, k = 1, \cdots, M$ that should represent the objects $O_i, i = 1, \cdots, N$ and use the following partial similarity measure [13] between object $i$ and concept $k$

$$l(i \mid k) = \left(2\pi\sigma_k^2\right)^{-1/2} \exp\left[-(O_i - S_k)^2/2\sigma_k^2\right] \qquad (1)$$

where, at this stage, the fuzziness $\sigma_k$ is a parameter given *a priori*. The goal is to find an assignment between models and objects such that the global similarity

$$L = \sum_i \log \sum_k l(i \mid k) \qquad (2)$$

is maximized. This is achieved by evolving the concept-models according to the dynamics

$$dS_k/dt = \sum_i f(k \mid i)\left[\partial \log l(i \mid k)/\partial S_k\right] \qquad (3)$$

where the fuzzy association variables $f(k \mid i)$ are defined by

$$f(k \mid i) = l(i \mid k) \Big/ \sum_{k'} l(i \mid k') \qquad (4)$$

and give a measure of the correspondence between object $i$ and concept $k$ relative to all other concepts $k'$. A salient feature of dynamic logic is a match between parameter uncertainty and fuzziness of similarity. In what follows we decrease the fuzziness during the time evolution of the modeling fields according to the prescription

$$\sigma_k^2(t) = \sigma_{k1}^2 \exp(-\alpha t) + \sigma_{k0}^2 \qquad (5)$$

with $\alpha = 5 \times 10^{-4}$, $\sigma_{k1} = 1 \forall k$ and $\sigma_{k0} = 0.03 \forall k$. In [16] we have shown that this setting allows perfect categorization, in a sense that the values of the modeling fields match those of the objects, provided that the number of modeling fields $M$ is equal or greater than the number of objects $N$. For $M = N$ there are $M!$ distinct but equally satisfactory assignments between concepts and objects. The initial conditions $S_k(t = 0)$ determine to which particular assignment the dynamics will converge.

We can easily be deceived by the apparent trivialness of this categorization task, since the categorization mechanisms built in our minds immediately sprout a one-to-one (if $N = M$) correspondence between objets and concepts. However, if asked to formalize that mechanism, the solutions proposed are usually very sophisticated, such as Steels' discrimination trees [3]. The key point in this task seems to be the symmetry-breaking of the permutation group associated to the labeling of objects by concepts. MFT provides an ingenious method to implement that partition in a fully autonomous framework. Moreover, the very same scheme used here which represents each objects by a point

on a single axis generalizes straightforwardly to the more realistic case in which the objects are represented by sets of points drawn from Gaussian distributions [17].

Fig. 1 illustrates the categorization mechanism in action for 8 objects, $O_i = i/10, i = 1, \ldots, 8$ and 8 concept-models $S_k, k = 1, \ldots, 8$ the initial values of which are chosen randomly. The modeling field dynamic equations (3) – (5) were solved numerically with Euler's method using the step-size $\varepsilon = 10^{-4}$. After convergence, a one-to-one mapping between objects and meanings is produced, namely, $O_1 = S_6$, $O_2 = S_1$, $O_3 = S_3, O_4 = S_4$, $O_5 = S_2$, $O_6 = S_5$, $O_7 = S_8$, and $O_8 = S_7$. The key point to our purposes is the interpretation of the index of the concept-model that becomes associated with a given object as the internal label of that object. This correspondence defines the permutation matrix $Q$, the nonzero entries of which indicate which meaning is assigned to which object. For example, $q_{ik} = 1$ indicates that, in presence of object $i$, the agent evokes meaning $k$. The transverse of this matrix is also important: $\left(q^T\right)_{ki} = 1$ indicates that, in the absence of external stimuli, the agent associates meaning k to object-model i. Hence distinct individuals, characterized by different initial values of the modeling fields, are likely to develop distinct labels for the same object, i.e., are characterized by different $Q$ matrices. The next section describes how communication can be established in such adverse situation.
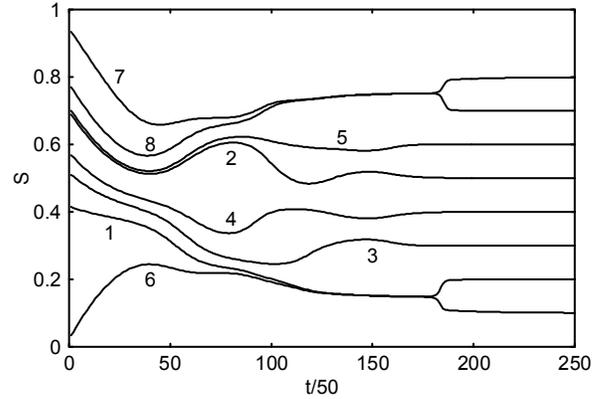


Fig. 1. Time evolution of the modeling fields for *M=N*=8 with randomly chosen initial conditions. The labels of the concept-models are indicated in the figure and the final one-to-one mapping object-concept is described in the text.

## II. EVOLVING COMMUNICATION

Following Smith [5], [6] we let the agents first to develop the meaning structure (i.e., the object-meaning mapping) and only then begin the communication phase. This procedure is in agreement with the admittedly arguable idea that the mental creation and manipulation of symbols came before communication. We assume that each agent, when communicates, can produce a signal for any of its concept-

models $k = 1, \ldots, M$. More specifically, the agents can choose any of $H$ different signals, which we denote by letters of the alphabet *a, b, c,* etc., to represent a concept. In doing so we are actually modeling the emergence of a holistic communication code, in which a signal stands for the meaning as a whole, so this formulation is more appropriate to study the emergence of protolanguage rather than of language [1]. At this stage, we can already point out a major difference between our approach and Smith's [5], [6]: we assume that with every meaning there is associated a, not necessarily distinct, signal, whereas Smith assumes that with every signal there is associated a, not necessarily different, meaning. As a result, in Smith's formulation there might be meanings without their corresponding signals, whereas in our case there might be signals without meaning, which seems a more reasonable working assumption.

Once produced, the signal is transferred from one agent – the signaler – to another agent – the receiver, which must interpret the signal from the context in which it is observed. At the beginning each agent has a different meaning-signal mapping, i.e., a lexicon of association between meaning and signals for use both in production and interpretation. Effective communication can take place provided the agents can reach a consensus on which signal must be assigned to each object (though there is no a priori mapping between object and signals). This consensus is usually achieved through language evolutionary games, in which the lexicon evolves from generation to generation guided by the increase of a payoff function, which essentially measures the communication accuracy of the population [18]-[22]. Here we take a culturally based view of language evolution and assume that the lexicons (or communication codes) are modified solely through learning.

For simplicity in this contribution we consider a population composed of two agents only, that play in turns the roles of signaler and receiver. Each agent is characterized by a $M \times H$ probability matrix $P$ whose entries $p_{kh} \in [0,1]$ yield the probability that meaning $k$ is associated with signal $h$. As mentioned before, we have $\sum_h p_{kh} = 1, \forall k$ in contrast to Smith's approach that introduces a similar quantity, except that the normalization is obtained by summing over all meanings $k$. We refer to $P$ as the verbalization matrix, since it describes completely the communicative behavior of the agents (see below). Learning consists in modifying the (initially random) matrix $P$ through an inference procedure based on the obverter scheme [14]. In the following we describe two learning procedures that differ basically on whether the agents receive feedback (supervised learning) or not (unsupervised learning) about the success of a communication episode.

### A. The unsupervised learning procedure

Two objects $i$ and $j$ are chosen randomly from the environment to form the context of the communication episode. The signaler, say agent *I*, picks randomly one of these objects, say *i*, retrieves its associated meaning, say *k*, and then emits a signal. This signal is chosen as the entry with the largest value in the row $p_{kh}^I$. Suppose the emitted signal is *a*. On the other side, the receiver, say agent *J*, which also has access to the context, must now interpret signal *a*. It does this in two steps. First, it finds which meanings are associated with the objects in the context, by looking at the entries of its matrix $Q^J$. Suppose it finds the correspondences $i \to l$ and $j \to m$. Second, it must decide which of these two meanings signal *a* is associated with. Since there is no additional information to make this choice, the learning procedure amplifies the entries $p_{la}^J$ and $p_{ma}^J$ by a factor $\alpha > 1$, so the new entries become $\alpha_l p_{la}^J$ and $\alpha_m p_{ma}^J$. As $P^J$ is a probability matrix the entries $p_{lh}^J, h \neq a$ and $p_{mh}^J, h \neq a$ must be reduced by the factor $\beta_k = \left(1 - \alpha_k p_{ka}^J\right)/\left(1 - p_{ka}^J\right)$ with $k = l, m$. To prevent $\beta_k$ to become negative we choose $\alpha_k = 1.01$ if $p_{ka}^J < 0.9$ and $\alpha_k = 0.99/p_{ka}^J$ otherwise, hence the need to identify the amplification factor by the meaning index $k = l, m$. This procedure can be interpreted as the lateral inhibition of the competing associations.

To proceed further, we must assume that the agents have a "Theory of Mind" (ToM) [23], i.e., that the receiver is somehow able to understand that the emitter thinks similar to itself and hence would behave likewise when facing the same situation. Accordingly, the receiver decides for the meaning that corresponds to the largest of the two entries $p_{la}^J$ and $p_{ma}^J$, i.e., it chooses the meaning that it itself would be most likely to associate with signal *a*. The original obverter scheme [14] assumes that the receiver has access to the verbalization matrix of the signaler (through mind-reading, as the critics were ready to point out) and so it chooses the meaning that corresponds to the largest of $p_{la}^I$ and $p_{ma}^I$, instead. Here we follow the more reasonable scheme, dubbed *introspective obverter* [6], which "solely" requires to endow the agents with a ToM rather than with telepathic abilities.

Finally, by using the transverse of the matrix $Q$ the object associated to the inferred meaning is retrieved. This finishes one learning episode that must be repeated very many times with each agent taking turn as signaler and receiver. Note that in this scheme only the receiver updates the verbalization matrix *P*. Communicative success is based on referent identity: signaler and receiver communicate successfully by referring to the same object, though they probably use different meanings to do so.

This learning scheme differs from that used by Smith in (at least) three aspects. First and as already said, our definition of the verbalization matrix guarantees that a concept can always be associated with a signal, though the reverse is not true. Second, the mechanism of amplification and inhibition of the entries of verbalization matrix described above dispenses the counter used to store the number of times each pair meaning-signal occurred. In addition, introduction of the matrix *P* makes our formulation similar to that employed in evolutionary language games, in which agents start with

random meaning-signal mappings [18]-[22]. Third, in our approach the verbalization matrix is updated only when the agent plays the role of receiver, whereas in Smith's approach also the entry corresponding to the meaning-signal picked up by the signaler ($p_{ka}^{I}$ in our example) is amplified.

An interesting feature of this learning scheme which, except for the points mentioned before is essentially the scheme used by Smith [6], [7] (see also [24] for an alternative learning algorithm), is that the agents receive no feedback about the success of their communication events: the modification of the verbalization matrix in context is the only way in which the agents learn. This is the reason we refer to it as unsupervised learning. The situation here is identical to the cross-situational learning scenario [25] in which the agents infer the meaning of a given word by monitoring its occurrence in a set of meanings. In this aspect, this scheme contrasts starkly with the procedure adopted by Steels [4], [5] described next.

## B.   The supervised learning procedure

The setting is identical to that described before except that the receiver must communicate its choice to the signaler (using some nonlinguistic means, such as pointing to the chosen object) and, in turn, the signaler must provide another nonlinguistic hint to indicate which object was the correct one in the context. Carrying on the example used to illustrate the unsupervised learning scheme, let us suppose first that the receiver decided for object $i$, which happens to be the correct choice. In this case, signaler and receiver amplify the entries $p_{ka}^{I}$ and $p_{la}^{J}$, respectively, using exactly the same procedure described in the unsupervised scheme, which includes the inhibition of the competing associations of the other signals with meaning $k$ in agent $I$ and with meaning $l$ in agent $J$. Next, suppose the receiver decided for object $j$, the wrong choice. In this case, both entries $p_{ka}^{I}$ and $p_{ma}^{J}$ are reduced by a factor $\gamma < 1$ (we set $\gamma = 0.95$), so that the new entries become $\gamma p_{ka}^{I}$ and $\gamma p_{ma}^{J}$. Simultaneously, all other signal associations with meanings $k$ must be amplified by the factor $\delta^{I} = (1 - \gamma p_{ka}^{I})/(1 - p_{ka}^{I})$, and similarly for meaning $m$ of agent $J$. Note that any choice of $\gamma < 1$ is sufficient to guarantee that $\delta^{I} > 1$. This is essentially the learning scheme used by Steels in the Talking Heads experiments [4], [5] (see also [26] for a detailed explanation of the learning algorithm).

The weak point of this learning scheme is the need for nonlinguistic hints to communicate the success or failure of the communication episode. This implies that, prior to learning, the agents are already capable to communicate (and understand) sophisticated meanings such as success and failure and behave (by updating their verbalization matrices) accordingly. In fact, as pointed out in [24], feedback about the outcome of the communication episode is a form of meaning transfer.

## C.   The Simulations

In the following we assume that agent $I$ is characterized by the object-meaning mapping $\langle o, m \rangle$ produced by the MFT dynamics illustrated in Fig. 1, namely, $\langle 1,6 \rangle^{I}$, $\langle 2,1 \rangle^{I}$, $\langle 3,3 \rangle^{I}$, $\langle 4,4 \rangle^{I}$, $\langle 5,2 \rangle^{I}$, $\langle 6,5 \rangle^{I}$, $\langle 7,8 \rangle^{I}$, and $\langle 8,7 \rangle^{I}$. The same procedure was used to generate the object-meaning mapping of agent $J$, but using different initial conditions for the modeling fields, resulting in the following mapping $\langle 1,3 \rangle^{J}$, $\langle 2,1 \rangle^{J}$, $\langle 3,6 \rangle^{J}$, $\langle 4,5 \rangle^{J}$, $\langle 5,8 \rangle^{J}$, $\langle 6,4 \rangle^{J}$, $\langle 7,7 \rangle^{J}$, and $\langle 8,2 \rangle^{J}$.
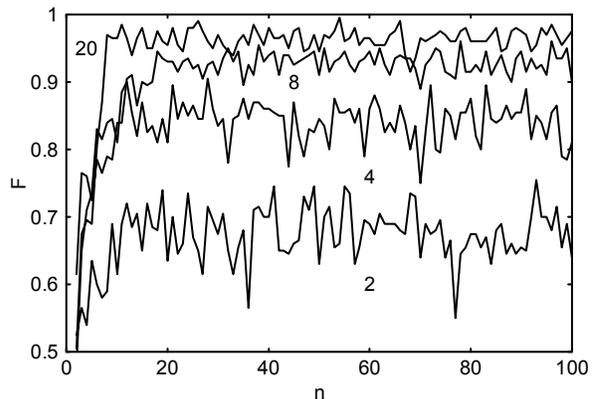


Fig. 2.  Fraction of successful communication events measured during the unsupervised learning procedure between interactions $(n-1)\Delta$ and $n\Delta$ with $\Delta = 100$. The alphabet size is $H = 2,4,8$, and 20 as indicated, and the number of objects and meaning is $N = M = 8$.

Here we focus mainly on the communication accuracy $F$ of the two agents $I$ and $J$, which is given by the fraction of successful communication events, i.e., events in which the receiver inferred correctly the object that the signaler had singled out from the context. To better illustrate the evolution of this quantity as the two agents interact, in Fig. 2 we measure $F$ for a fixed number of interactions $\Delta = 100$ as the unsupervised learning proceeds. We define one interaction as two sequential communication episodes, allowing thus the agents to take turns as signaler and receiver. The integer $n = 1,2,\dots$ in the X-axis of this graph indicates that $F$ was measured between interactions $(n-1)\Delta$ and $n\Delta$. Clearly, since the context comprises two objects only, pure guessing yields $F \approx 0.5$. Interestingly, although the agents have a set of $H$ distinct signals to choose from, they do not use the entire repertoire of signals. For example, in the case $H = 8$ illustrated in Fig. 2, the agents actually use only 5 distinct signals. Nevertheless, this yields the average fraction of successes $\langle F \rangle \approx 0.89$, which corresponds to a very good performance. In fact, the accuracy loss of using the same signal for different objects is small because usually the context eliminates the ambiguity. In the unsupervised scheme, failure may occur only when these objects make up the context, but even then there is a 50% chance of correct guessing.

To be more quantitative we run 1000 replicates of the experiment depicted in Fig. 2 for both the unsupervised and supervised learning procedures. For each replicate we measure the communication accuracy in the stationary regime performing an average over the last 1000 interactions. In addition, we measure the effective number of signals $H'$ used by the agents after their communication codes become fixed. These data are then averaged over all replicates and the result shown in Figs. 3 and 4. Inspection of these figures makes it clear that the unsupervised learning scheme failed to produce the maximum communication accuracy because the agents actually used fewer signals than the necessary to generate a one-to-one correspondence between signals and meanings (or objects). These ideal codes can be obtained by the supervised learning procedure provided the size $H$ of the repertoire of signals available to the agents is sufficiently large. It is interesting that the agents *never* use more signals than the number of meanings, which would amount to assign different signals to the same meaning. This phenomenon, known as synonymy, is very rare in language (it is hard to find two words that have exactly the same meaning) and it seems to be automatically ruled out by the two learning procedures used in our simulations. In addition, we note from Fig. 4 that $H'$ increases linearly with $H$ for $H < M = 8$ and then begins to level off at some value that depends on the learning scheme ($H' \to 5.5$ for the unsupervised and $H' \to 8$ for the supervised scheme).

We have found that, similarly to the language evolutionary games [21], both learning algorithms lead always to binary verbalization matrices $P$, i.e., matrices whose entries $p_{kh}$ can take on the values 0 or 1 only. Together with the constraint $\sum_h p_{kh} = 1, \forall k$, this observation excludes the possibility of synonymy altogether, since if $p_{ka} = 1$ then $p_{kh} = 0$ for $h \neq a$. Homonymy (i.e., signals that have more than one meaning), however, is an absorbing state of the learning procedures and seems to be the usual outcome of both learning schemes. Since these procedures may be viewed as algorithms to maximize the communication accuracy, the verbalization matrices associated to homonymy can then be interpreted as local maxima of the learning dynamics. The enormous difficulty to reach a global maxima (i.e., a one-to-one signal-meaning correspondence) illustrated in Fig. 3 has recently been reported in the context of evolutionary language games as well [22]. The interesting finding here is that the increase of the size of the repertoire of signals allows the supervised learning scheme to escape the local maxima and ultimately reach an optimal communication code.
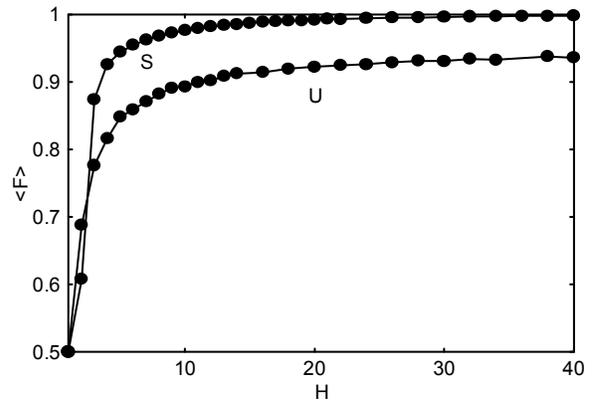


Fig. 3. Average communication accuracy at the stationary state, obtained with 1000 replicates of the experiment shown in Fig. 2 for both the unsupervised (U) and the supervised (S) learning schemes, as function of the alphabet size $H$ for $N = 8$ objects and $M = 8$ concepts. The agents do not use all the available repertoire of signals as shown in Fig. 4.
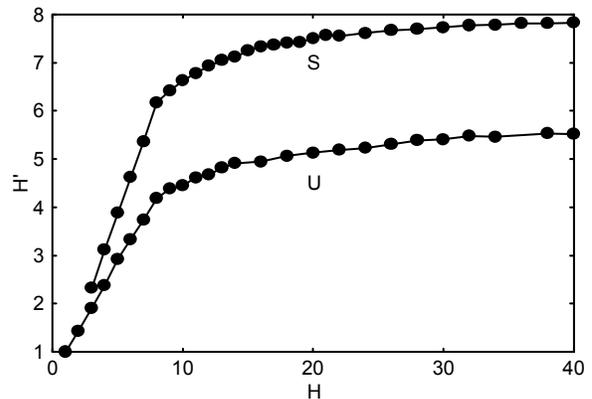


Fig. 4. Average number of signals used by the agents in the experiment illustrated in Fig. 3 as function of the alphabet size $H$. To produce perfect communication the agents should use the same number of signals as objects, $N = 8$ in this case. Failure to achieve that in the case of unsupervised learning prevents successful communication. .

## III. CONCLUSION

This work represents a modest first step to tackle the fundamental problem of the co-evolution of language and cognition. In the present setting we have considered the case that the agents develop (usually) different meanings for the same object - this guarantees that mind-reading does not occur since it would simply be useless. But at this stage we do not consider the possibility that signals can create novel meanings, which are not directly grounded to objects in the agent's world. (Of course, these meanings must necessarily be combinations of the grounded ones, see, e.g., [1], [10].) Before addressing this complex situation, however, we plan to consider in a future publication a simpler scenario for the co-evolution of language and cognition. The key issue is to include in the model the option that the object-meaning mapping be modified (or expanded) as a result of the labeling of meanings with words, i.e., of naming the objects. In fact, preliminary results indicate that naming can greatly improve the differentiation capability of the agents [27].

In summary, our study of the performance of the unsupervised learning procedure inspired on the work of Smith [6], [7] indicates that the unsupervised scheme fails to produce ideal communication codes, i.e., codes that implement a one-to-one correspondence between meanings and signals. We note that this conclusion was reached in a best case scenario in which the context comprised two objects and the population two agents only. On the other hand, the supervised learning scheme, based on the proposal by Steels and Kaplan [5], [26], does succeed in that task, provided the size of the repertoire of signals is set to a large value, although synonymy is never observed. In other words, the agents must be capable to generate and choose among tens of distinct sounds to associate with a given meaning in order to be able to produce an ideal communication code, which actually uses a few signals only. This is a very odd finding: the mere ability to produce different signals bears on the quality of the evolved communication codes. Nevertheless, the supervised learning scheme is not entirely satisfactory from the perspective of the evolution of language or protolanguage since it presupposes that the agents are a priori capable of exchanging information about the success of their communication episodes, i.e., it assumes some form of meaning transfer [24], [25]. An alternative framework to evolve ideal communication codes, that dispenses altogether with the assumption of nonlinguistic means to exchange highly relevant information, is the language evolutionary games in which communication success is directly tied to the survival and reproduction of the agents [18]-[22]. Hence natural selection takes care of informing the agents of their success or failure in the communication game.

## REFERENCES

[1] D. Bickerton, *Language & Species*. University of Chicago Press, Chicago, 1990.

[2] D. Bickerton, *Language and Human Behavior*. University of Washington Press, Seattle, 1995.

[3] L. Steels, "Perceptually grounded meaning creation," In: M. Tokoro (Ed.), ICMAS-96 *Proceedings Second International Conference on Multi-Agent Systems*. AAAI Press, Menlo Park, pp. 338-344, 1996.

[4] L. Steels, "Grounding symbols through evolutionary language games," In: A. Cangelosi, D. Parisi (Eds.), *Simulating the Evolution of Language*, London: Springer-Verlag, pp. 211-226, 2002.

[5] L. Steels and F. Kaplan, "Situated grounded word semantics," *Proceedings of IJCAI-99*. Morgan Kauffman, Los Angeles, pp. 862-867, 1998.

[6] A. D. M. Smith, "Semantic generalization and the inference of meaning," In: W. Banzhaf, T. Christaller, P. Dittrich, J. T. Kim, J. Ziegler (Eds.), *Proceedings 7th European Conference on Artificial Life*, Lecture Notes in Artificial Intelligence, vol. 2801, pp. 499-506, 2003.

[7] A. D. M. Smith, "Intelligent meaning creation in a clumpy world helps communication", *Artificial Life*, vol. 9, pp. 557-574, 2003.

[8] T. W. Deacon, *The Symbolic Species*. W. W. Norton & Company, New York, 1997.

[9] L. Steels, "How language bootstraps cognition," In: I. Wachsmutt, B. Jung (Eds.), *Proceedings der 4 Fachtagung der Gesellschaft fuer Kognitionswissenschaft*. Infix, Sankt Augusting, pp. 1-3, 1999.

[10] A. Cangelosi, A. Greco, and S. Harnard, "Symbol grounding and the symbolic theft hypothesis," In: A. Cangelosi, D. Parisi (Eds.), *Simulating the Evolution of Language*, London: Springer-Verlag, pp. 191-210, 2002.

[11] L. I. Perlovsky, "Integrating language and cognition", *IEEE Connections*, vol. 2, no. 2, pp. 8-13, May 2004.

[12] C. L. Nehaniv, "The making of meaning in societies: semiotic & information –theoretic background to the evolution of communication," *Proceedings of the AISB Symposium: Starting from Society – the application of social analogies to computational systems*, pp. 73-84, 2000.

[13] L. I. Perlovsky, *Neural Networks and Intellect: Using Model-Based Concepts*. Oxford University Press , Oxford, 2001.

[14] M. Oliphant and J. Batali, "Learning and the emergence of coordinated communication," *Center for Research on Language Newsletter*, vol 1, 1997.

[15] L. Steels and T. Belpaeme, "Coordinating perceptually grounded categories through language. A case study for color", *Behavioral and Brain Sciences*, vol. 24, no. 8, pp. 469-529, 2005.

[16] J. F. Fontanari and L. I. Perlovsky, "Meaning creation and modeling field theory," In: C. Thompson, H. Hexmoor (Eds.), *Proceedings International Conference on Integration of Knowledge Intensive Multi-Agent Systems*. IEEE Press, Boston, pp. 405-410, 2005.

[17] J. F. Fontanari and L. I. Perlovsky, "Categorization and symbol grounding in a complex environment" *Proceedings of IJCNN2006: 2006 International Joint Conference on Neural Networks*. Vancouver, July 2006.

[18] J. R. Hurford, "Biological evolution of the Saussurean sign as a component of the language acquisition device," *Lingua*, vol. 77, pp. 187-222, 1989.

[19] M. Oliphant, "The dilemma of Saussurean communication," *BioSystems*, vol. 37, pp. 31-38, 1996.

[20] M. A. Nowak and D. C. Krakauer, "The evolution of language," *Proc. Natl. Acad. Sci. USA*, vol. 96, pp. 8028-8033, 1999.

[21] M. A. Nowak, J. B. Plotkin, and D. C. Krakauer, "The Evolutionary Language Game," *J. theor. Biol.*, vol. 200, pp. 147-162, 1999.

[22] J. F. Fontanari and L. I. Perlovsky, "Evolving compositionality in evolutionary language games," *Trends Evol. Comp.*, submitted for publication.

[23] M. Donald, *Origins of the Modern Mind*. Harvard University Press, Cambridge, 1991.

[24] J. de Beule, B. de Vylder, and T. Belpaeme, "A cross-situational learning algorithm for damping homonymy in the guessing game," In: L.M. Rocha, M. Bedau, D. Floreano, R. Goldstone, A. Vespignani, L. Yaeger (Eds.), *Proceedings of the Xth Conference on Artificial Life*, The MIT Press, Cambridge, 2006.

[25] J. M. Siskind, "A computational study of cross-situational techniques for learning word-to-meaning mappings," *Cognition*, vol. 61, pp.39-91, 1996.

[26] L. Steels and F. Kaplan, "Spontaneous lexicon change," *Proceedings of COLING-ACL*. Morgan Kaufmann, San Francisco, pp. 1243-1250, 1998.

[27] L. I. Perlovsky and J. F. Fontanari, "How language can guide intelligence", *Proceedings of the Sixth International Conference on the Evolution of Language*. Rome, April 2006.

# Language and Cognition Integration through Modeling Field Theory: Category Formation for Symbol Grounding

Vadim Tikhanoff [(1)], José F. Fontanari [(2)], Angelo Cangelosi [(1)],
Leonid I. Perlovsky [(3)]

(1)   Adaptive Behaviour & Cognition, University of Plymouth, Plymouth PL4 8AA, UK
(2) Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, SP, Brazil
(3) Air Force Research Laboratory, Hanscom Air Force Base, MA 01731, USA
vadim.tikhanoff@plymouth.ac.uk; fontanari@ifsc.usp.br; acangelosi@plymouth.ac.uk;
leonid.perlovsky@hanscom.af.mil

Neural Modeling Field Theory is based on the principle of associating lower-level signals (e.g., inputs, bottom-up signals) with higher-level concept-models (e.g. internal representations, categories/concepts, top-down signals) avoiding the combinatorial complexity inherent to such a task. In this paper we present an extension of the Modeling Field Theory neural network for the classification of objects. Simulations show that (i) the system is able to dynamically adapt when an additional feature is introduced during learning, (ii) that this algorithm can be applied to the classification of action patterns in the context of cognitive robotics and (iii) that it is able to classify multi-feature objects from complex stimulus set. The use of Modeling Field Theory for studying the integration of language and cognition in robots is discussed.

## Introduction

### Grounding language in categorical representations

A growing amount of research on interactive intelligent systems and cognitive robotics is focusing on the close integration of language and other cognitive capabilities [1,3,13]. One of the most important aspects in language and cognition integration is the grounding of language in perception and action. This is based on the principle that cognitive agents and robots learn to name entities, individuals and states in the external (and internal) world whilst they interact with their environment and build sensorimotor representations of it. For example, the strict relationship between language and action has been demonstrated in various empirical and theoretical studies, such as psycholinguistic experiments [10], neuroscientific studies [16] and language evolution theories [17]. This link has also been demonstrated in computational models of language [5,21].

Approaches based on language and cognition integration are based on the principle of grounding symbols (e.g. words) in internal meaning representations. These are

normally based on categorical representations [11]. Much research has been dedicated on modeling the acquisition of categorical representation for the grounding of symbols and language. For example, Steels [19,20] has studied the emergence of shared languages in group of autonomous cognitive robotics that learn categories of objects. He uses discrimination tree techniques to represent the formation of categories of geometric shapes and colors. Cangelosi and collaborators have studied the emergence of language in multi-agent systems performing navigation and foraging tasks [2], and object manipulation tasks [6,12]. They use neural networks that acquire, through evolutionary learning, categorical representations of the objects in the world that they have to recognize and name.


**Modeling Field Theory**

Current grounded agent and robotic approaches have their own limitations. For example, one important issue is the scaling up of the agents' lexicon. Present models can typically deal with a few tens of words (e.g. [20]) and with a limited set of syntactic categories (e.g. nouns and verbs in [2]). This is mostly due to the use of computational intelligent techniques, the performance of which is considerably degraded by the combinatorial complexity (CC) of this problem. The issue of scaling up and combinatorial complexity in cognitive systems has been recently addressed by Perlovsky [14]. In linguistic systems, CC refers to the hierarchical combinations of bottom-up perceptual and linguistic signals and top-down internal concept-models of objects, scenes and other complex meanings. Perlovsky proposed the neural Modeling Field Theory (MFT) as a new method for overcoming the exponential growth of combinatorial complexity in the computational intelligent techniques traditionally used in cognitive systems design. Perlovsky [15] has suggested the use of MFT specifically to model linguistic abilities. By using concept-models with multiple sensorimotor modalities, a MFT system can integrate language-specific signals with other internal cognitive representations.

Modeling Field Theory is based on the principle of associating lower-level signals (e.g., inputs, bottom-up signals) with higher-level concept-models (e.g. internal representations, categories/concepts, top-down signals) avoiding the combinatorial complexity inherent to such a task. This is achieved by using measures of similarity between concept-models and input signals together with a new type of logic, so-called dynamic logic. MFT may be viewed as an unsupervised learning algorithm whereby a series of concept-models adapt to the features of the input stimuli via gradual adjustment dependent on the fuzzy similarity measures.

A MFT neural architecture was described in [14]. It combines neural architecture with models of objects. For feature-based object classification considered here, each input neuron $i = 1, \ldots, N$ encodes feature values $O_i$ (potentially a vector of several features); each neuron $i$ may contain a signal from a real object or from irrelevant context, clutter, or noise. We term the set $O_i, i = 1, \ldots, N$ an input neural field: it is a set of bottom-up input signals. Top-down, or priming signal-fields to these neurons are generated by models, $M_k(S_k)$ where we enumerate models by index $k = 1, \ldots, M$. Each model is characterized by its parameters $S_k$, which may also be a vector of

several features. In this contribution we will consider the simplest possible case, in which parameters model represent feature values of object, $M_k(S_k) = S_k$. Interaction between bottom-up and top-down signals is determined by neural weights associating signals and models as follows. We introduce an arbitrary similarity measure $l(i\,|\,k)$ between bottom-up signals $O_i$ and top-down signals $S_k$ [see equation (2)], and define the neural weights by

$$f(k\,|\,i) = l(i\,|\,k)\Big/\sum_{k'} l(i\,|\,k') \,. \tag{1}$$

These weights are functions of the model parameters $S_k$, which in turn are dynamically adjusted so as to maximize the overall similarity between object and models. This formulation sets MFT apart from many other neural networks.

Recently, MFT has been applied to the problem of categorization and symbol grounding in language evolution models. Fontanari and Perlovsky [7] use MFT as an alternative categorization and meaning creation method to that of discrimination trees used by Steels [19]. They consider a simple world composed of few objects characterized by real-valued features. Whilst in Steels's work each object is defined by 9 features (e.g. vertical position, horizontal, R, G and B color component values), here each object consists of a real-valued number that identifies only one feature (sensor). The task of the MFT learning algorithm is to find the concept-models that best match these values. Systematic simulations with various numbers of objects, concept-models and object/model ratios, show that the algorithm can easily learn the appropriate categorical model. This MFT model has been recently extended to study the dynamic generation of concept-models to match the correct number of distinct objects in a complex environment [8]. They use the Akaike Information Criterion to gradually add concept-models until the system settles to the correct number of concepts, which corresponds to the original number of distinct objects defined by the experimenter. This method has been applied to complex classification tasks with high degree of variance and overlap between categories. Fontanari and Perlovsky [9] have also used MFT in simulations on the emergence of communication. Meanings are created through MFT categorization, and word-meaning associations are learned using two variants of the obverter procedure [18], in which the agents may, or may not, receive feedback about the success of the communication episodes. They show that optimal communication success is guaranteed in the supervised scheme, provided the size of the repertoire of signals is sufficiently large, though only a few signals are actually used in the final lexicon.

**MFT for categorization of multi-dimensional object feature representations**

The above studies have demonstrated the feasibility of using MFT to model symbol grounding and fuzzy similarity-based category learning. However, the model has been applied to a very simplified definition of objects, each consisting of one feature. Simulations have also been applied to a limited number of categories (concept-models). In more realistic contexts, perceptual representations of objects

consist of multiple features or complex models for each sensor, or result from the integration of different sensors. For example, in the context of interactive intelligent systems able to integrate language and cognition, their visual input would consist of objects with a high number of dimensions or complex models. These could be low-level vision features (e.g. individual pixel intensities), or some intermediate image processing features (e.g. edges and regions), or higher-level object features (color, shape, size etc.). In the context of action perception and imitation, a robot would have to integrate various input features from the posture of the teacher robot to identify the action or complex models (e.g. [6]). The same need for multiple-feature objects applies to audio stimuli related to language/speech. In addition, the interactive robot would have to deal with hundreds, or thousands, categories, and with high degrees of overlap between categories.

To address the issue of multi-feature representation of objects and that of the scaling up of the model we have extended the MFT algorithm to work with multiple-feature objects. We consider both the cases in which all features are present from the start, and the case in which the features are dynamically added during learning. For didactic purposes, first we will carry out simulations on very simple data sets, and then on data related to the problem of action recognition in interactive robots. Finally, we will present some results on the scale up of the model, using hundred of objects.

## The Model

We consider the problem of categorizing $N$ objects $i = 1, \cdots, N$, each of which characterized by $d$ features $e = 1, \cdots, d$. These features are represented by real numbers $O_{ie} \in (0,1)$ - the input signals - as described before. Accordingly, we assume that there are $M$ $d$-dimensional concept-models $k = 1, \cdots, M$ described by real-valued fields $S_{ke}$, with $e = 1, \cdots, d$ as before, that should match the object features $O_{ie}$. Since each feature represents a different property of the object as, for instance, color, smell, texture, height, etc. and each concept-model component is associated to a sensor sensitive to only one of those properties, we must, of course, seek for matches between the same component of objects and concept-models. Hence it is natural to define the following partial similarity measure between object $i$ and concept $k$

$$l(i \mid k) = \prod_{e=1}^{d} \left(2\pi\sigma_{ke}^2\right)^{-1/2} \exp\left[-\left(S_{ke} - O_{ie}\right)^2 / 2\sigma_{ke}^2\right] \tag{2}$$

where, at this stage, the fuzziness $\sigma_{ke}$ is a parameter given *a priori*. The goal is to find an assignment between models and objects such that the global similarity

$$L = \sum_i \log \sum_k l(i \mid k) \tag{3}$$

is maximized. This maximization can be achieved using the MFT mechanism of concept formation which is based on the following dynamics for the modeling field components

$$dS_{ke}/dt = \sum_i f(k\,|\,i)\left[\partial \log l(i\,|\,k)/\partial S_{ke}\right], \qquad (4)$$

which, using the similarity (1), becomes

$$dS_{ke}/dt = -\sum_i f(k\,|\,i)\left(S_{ke} - O_{ie}\right)/\sigma_{ke}^2 \; . \qquad (5)$$

Here the fuzzy association variables $f(k\,|\,i)$ are the neural weights defined in equation (1) and give a measure of the correspondence between object $i$ and concept $k$ relative to all other concepts $k'$. These fuzzy associations are responsible for the coupling of the equations for the different modeling fields and, even more importantly for our purposes, for the coupling of the distinct components of a same field. In this sense, the categorization of multi-dimensional objects is not a straightforward extension of the one-dimensional case because new dimensions should be associated with the appropriate models. This nontrivial interplay between the field components will become clearer in the discussion of the simulation results.

It can be shown that the dynamics (4) always converges to a (possibly local) maximum of the similarity $L$ [14], but by properly adjusting the fuzziness $\sigma_{ke}$ the global maximum often can be attained. A salient feature of dynamic logic is a match between parameter uncertainty and fuzziness of similarity. In what follows we decrease the fuzziness during the time evolution of the modeling fields according to the following prescription

$$\sigma_{ke}^2(t) = \sigma_a^2 \exp(-\alpha t) + \sigma_b^2 \qquad (6)$$

with $\alpha = 5 \times 10^{-4}$, $\sigma_a = 1$ and $\sigma_b = 0.03$. Unless stated otherwise, these are the parameters we will use in the forthcoming analysis.

## Simulations

In this section we will report results from three simulations. The first will use very simple data sets that necessitate the use of two features to correctly classify the input objects. We will demonstrate the gradual formation of appropriate concept-models though the dynamic introduction of features. In the second simulation we will demonstrate the application of the multi-feature MFT on data related to the classification of actions from interactive robotics study. Finally, in the third simulation we will consider the scaling up of the MFT to complex data sets.

To facilitate the presentation of the results, we will interpret both the object feature values and the modeling fields as $d$-dimensional vectors and follow the time evolution of the corresponding vector length

$$S_k = \sqrt{\sum_{e=1}^{d} (S_{ke})^2 \Big/ d} \; , \qquad (7)$$

which should then match the object length $O_i = \sqrt{\sum_{e=1}^{d} (O_{ie})^2 \Big/ d}$ .

**Simulation I: Incremental addition of feature**

Consider the case in which we have the 5 objects, initially with only one-feature information. For instance, we can consider color information only on Red, the first of the 3 RGB feature values, as used in Steels's [19] discrimination-tree implementation. The objects have the following R feature values: $O_1 = [0.1]$, $O_2 = [0.2]$, $O_3 = [0.3]$, $O_4 = [0.5]$, $O_5 = [0.5]$.

A first look at the data indicates that these 5 input stimuli belong to four color categories (concept-models) with Red values respectively 0.1, 0.2, 0.3 and 0.5. As a matter of fact, the application of the MFT algorithm to the above mono-dimensional input objects reveal the formation of 4 model fields, even when we start with the condition in which 5 fields are randomly initialized (Fig. 1).
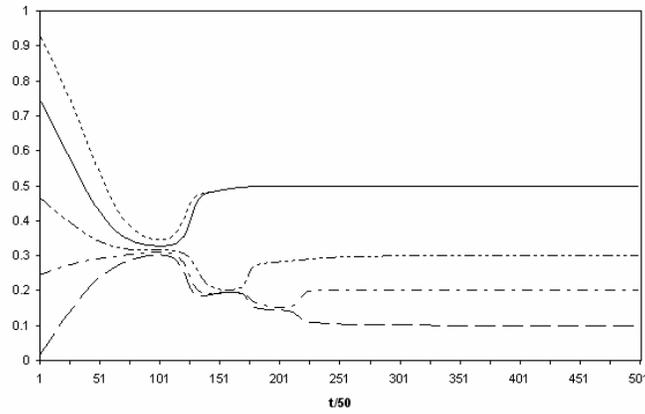


Fig. 1 – Time evolution of the fields with only the first feature being used as input. Only 4 models are found, with two initial random fields converging towards the same .5 Red concept-model value.

Let us now consider the case in which we add information from the second color sensor, Green. The object input data will now look like these: $O_1 = [0.1, 0.4]$, $O_2 = [0.2, 0.5]$, $O_3 = [0.3, 0.2]$, $O_4 = [0.5, 0.3]$, $O_5 = [0.5, 0.1]$.

The same MFT algorithm is applied with 5 initial random fields. For the first 12500 training cycles (half of the previous training time), only the first feature is utilized. At timestep 12500, both features are considered when computing the fuzzy similarities. From timestep 12500, the dynamics of the $\sigma_2$ fuzziness value is initialized, following equation (7), whilst $\sigma_1$ continues[1] its decrease pattern started at timestep 0. Results in Fig. 2 show that the model is now able to correctly identify 5 different fields, one per combined RG color type.

---

[1] We have also experimented with the alternative method of re-initializing both $\sigma_e$ values, as in equation (7), whenever a new feature is added. This method produces similar results.
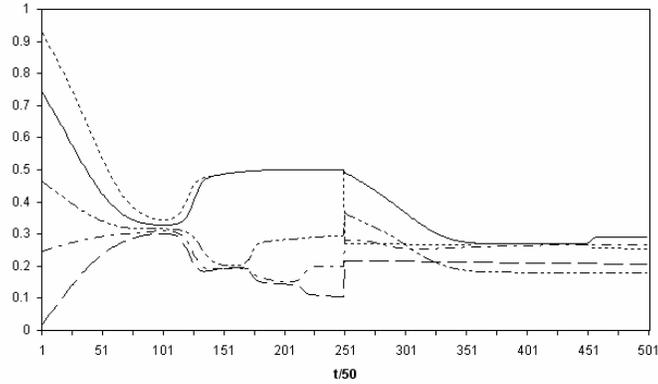
Fig. 2 – Time evolution of the fields when the second feature is added at timestep 12500. The dynamic fuzziness reduction for $\sigma_2$ starts at the moment the 2nd feature is introduced, and is independent from $\sigma_1$. Note the restructuring of 4 fields initially found up to timestep 12500, and the further discovery of the model. The fields values in the first 12500 cycles is the actual mono-dimensional field value, whilst from timestep 12500 the equation in (7) is used to plot the combined fields' value.
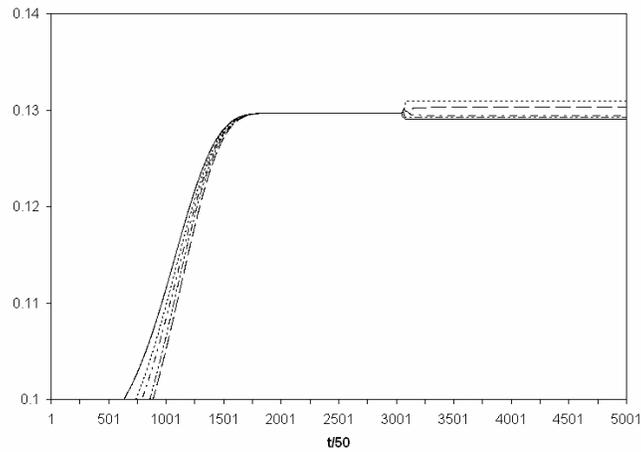


Fig. 3 – Evolution of fields in the robot posture classification task. The value of the field corresponds to equation (7). Although the five fields look very close, in reality the individual field values match very well the 42 parameters of the original positions.

## Simulation II: Categorization of robotic actions

In the introduction we have proposed the use of MFT for modeling the integration of language and cognition in cognitive robotic studies. This is a domain where the input to the cognitive agent (e.g. visual and auditory input) typically consists of multi-

dimensional data such as images of objects/robots and speech signals. Here we apply the multi-dimensional MFT algorithm to the data on the classification of the posture of robots, as in an imitation task. We use data from a cognitive robotic model of symbol grounding [4,6]. We have collected data on the posture of robots using 42 features. This consist of the 7 main data (X, Y, Z, and rotations of joints 1, 2, 3, and 4) for each of the 6 segments of the robot's arms (right shoulder, right upperarm, right elbow, left shoulder, left upperarm, left elbow). As training set we consider 5 postures: resting position with both arms open, left arm in front, right arm in front, both arms in front, and both arms down. In this simulation, all 42 features are present from timestep 0. Fig. 3 reports the evolution of fields and the successful identification of the 5 postures.
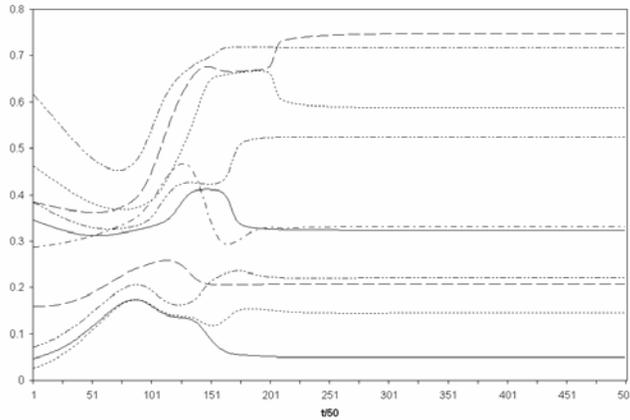


Fig. 4 – Evolution of fields in the case with 1000 input objects and 10 prototypes.

**Simulation III: Scaling up with complex stimuli sets**

Finally, we have tested the scaling-up of the multi-dimensional MFT algorithm with a complex categorization data set. The training environment is composed of 1000 objects belonging to the following 10 2-feature object prototypes: [0.1, 0.8], [0.2, 1.0], [0.3, 0.1], [0.4, 0.5], [0.5, 0.2], [0.6, 0.3], [0.7, 0.4], [0.8, 0.9], [0.9, 0.6] and [1.0, 0.7]. For each prototype, we generated 100 objects using a Gaussian distribution with standard deviation of 0.05. During training, we used 10 initial random fields.

Fig. 4 reports the time evolution of the 10 concept-models fields. The analysis of results also shows the successful identification of the 10 prototype models and the matching between the 100 stimuli generated by each object and the final values of the fields.

## Discussion and Conclusion

In this paper we have presented an extension of the MFT algorithm for the classification of objects. In particular we have focused on the introduction of multi-dimensional features for the representation of objects. The various simulations showed that (i) the system is able to dynamically adapt when an additional feature is introduced during learning, (ii) that this algorithm can be applied to the classification of action patterns in the context of cognitive robotics and (iii) that it is able to classify multi-feature objects from complex stimulus set.

Our main interest in the adaptation of MFT to multi-dimensional objects is for its use in the integration of cognitive and linguistic abilities in cognitive robotics. MFT permits the easy integration of low-level models and objects to form higher-order concepts. This is the case of language, which is characterized by the hierarchical organization of underlying cognitive models. For example, the acquisition of the concept of "word" in a robot consists in the creation of a higher-order model that combines a semantic representation of an object model (e.g. prototype) and the phonetic representation of its lexical entry [15]. The grounding of language into categorical representation constitutes a cognitively-plausible approach to the symbol grounding problem [11]. In addition, MFT permits us to deal with the problem of combinatorial complexity, typical of models dealing with symbolic and linguistic representation. Current cognitive robotics model of language typically deal with few tens or hundred of words (e.g. [6,19]). With the integration of MFT and robotics experiments we hope to deal satisfactory with the combinatorial complexity problem.

Ongoing research is investigating the use of MFT for the acquisition of language in cognitive robotics. In particular we are currently looking at the use of multi-dimensional MFT to study the emergence of shared languages in a population of robots. Agents first develop an ability to categorize objects and actions by building concept-models of objects prototypes. Subsequently, they start to learn a lexicon to describe these objects/actions through a process of cultural learning. This is based on the acquisition of a higher-order MFT.

## References

[1] Barsalou, L. (1999), Perceptual symbol systems. *Behavioral and Brain Sciences*, 22, 577-609.
[2] Cangelosi A. (2001). Evolution of communication and language using signals, symbols and words. *IEEE Transactions on Evolutionary Computation*. 5(2), 93-101

[3]    Cangelosi A., Bugmann G. & Borisyuk R. (Eds.) (2005). *Modeling Language, Cognition and Action: Proceedings of the 9th Neural Computation and Psychology Workshop*. Singapore: World Scientific.

[4]    Cangelosi A., Hourdakis E. & Tikhanoff V. (2006). Language acquisition and symbol grounding transfer with neural networks and cognitive robots. *Proceedings of IJCNN2006: 2006 International Joint Conference on Neural Networks*. Vancouver, July 2006.

[5]    Cangelosi, A., & Parisi, D. (2004). The processing of verbs and nouns in neural networks: Insights from synthetic brain imaging. *Brain and Language*, 89(2), 401-408.

[6]    Cangelosi A, Riga T (2006). An Embodied Model for Sensorimotor Grounding and Grounding Transfer: Experiments with Epigenetic Robots, *Cognitive Science*, 30(4), 1-17.

[7]    Fontanari J.F., Perlovsky L.I. (2005). Meaning creation and modeling field theory. In C. Thompson & H. Hexmoor (Eds.), *IEEE KIMAS2005: International Conference on Integration of Knowledge Intensive Multi-Agent Systems*. IEEE Press, pp. 405-410.

[8]    Fontanari J.F., Perlovsky L.I. (2006a). Categorization and symbol grounding in a complex environment. *Proceedings of IJCNN2006: 2006 International Joint Conference on Neural Networks*. Vancouver, July 2006.

[9]    Fontanari J.F., Perlovsky L.I. (2006b). Meaning creation and communication in a community of agents. *Proceedings of IJCNN2006: 2006 International Joint Conference on Neural Networks*. Vancouver, July 2006.

[10]   Glenberg A., & Kaschak, M. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, 9(3), 558-565.

[11]   Harnad, S. (1990). The symbol grounding problem. *Physica D*, 42, 335-346.

[12]   Marocco, D., Cangelosi, A., & Nolfi, S. (2003). The emergence of communication in evolutionary robots. *Philosophical Transactions of the Royal Society of London – A* 361, 2397-2421.

[13]   Pecher, D., & Zwaan, R.A., (Eds.). (2005). Grounding cognition: The role of perception and action in memory, language, and thinking. Cambridge: Cambridge University Press.

[14]   Perlovsky L. *Neural Networks and Intellect: Using Model-Based Concepts*. Oxford University Press, New York, 2001.

[15]   Perlovsky L., "Integrating language and cognition," *IEEE Connections*, vol. 2, pp. 8-13, 2004

[16]   Pulvermuller F. (2003) *The neuroscience of language. On brain circuits of words and serial order*. Cambridge: Cambridge University Press.

[17]   Rizzolatti, G., & Arbib, M. (1998). Language within our grasp. *Trends in Neuroscience*, 21: 188-194.

[18]   Smith A. D. M. (2003). Semantic generalization and the inference of meaning," In: W. Banzhaf, T. Christaller, P. Dittrich, J. T. Kim, J. Ziegler (Eds.), *Proceedings of the 7th European Conference on Artificial Life*, Lecture Notes in Artificial Intelligence, vol. 2801, pp. 499-506.

[19]   Steels L (1999) *The talking heads experiment (Volume I. Words and meanings)*. Antwerpen: Laboratorium.

[20]   Steels, L. (2003) Evolving grounded communication for robots. *Trends in Cognitive Sciences*, 7(7):308-312.

[21]   Wermter, S., Elshaw, M., and Farrand, S., 2003, A modular approach to self-organization of robot control based on language instruction. *Connection Science*, 15(2-3): 73-94.

# Statistical analysis of discrimination games

J. F. Fontanari[1]

[1]*Instituto de Física de São Carlos, Universidade de São Paulo,*
*Caixa Postal 369, 13560-970 São Carlos SP, Brazil*

The hypothesis that meanings originate from discrimination tasks, in which an individual attempts to categorize $N$ objects using a set of $M$ sensory channels, is examined within a quantitative statistical perspective. Failure in discrimination triggers the refinement of a randomly-chosen sensory channel, starting thus an ongoing process, termed discrimination game, that ends only when all objects are differentiated. We show that the expected number of trials of a discrimination game diverges in the case of a single channel and scales with the power $N^{2/M}$ for $M \geq 2$.

Any theory that purports to explain the evolution of language (or, more generally, of communication) must assume that the individuals are endowed with some innate categorization mechanism, which makes them capable of classifying different types of situations and, accordingly, of recognizing when a situation of a particular type turns up. Meanings express patterns of categorization, but are not innate. Rather, they are produced afresh in each individual, who creates a particular system of meanings based on its experiences [1]. Although the meaning of a given word is usually defined by its relationship to other words, and in terms of other words, at least a few words must be grounded in reality, so they can be used to identify actions and objects in the real world [2]. Since the groundbreaking work of de Saussure [3] it is known that words refer to real-world objects only indirectly as first the sense perceptions are mapped onto a conceptual representation - the meaning - and then this conceptual representation is mapped onto a linguistic representation - the words. Hence the need to taking into account mechanisms for perceptually grounded meaning creation in modeling language evolution.

Perceptually grounded meaning creation, viewed here as synonymous to category creation, underlies the current effort to develop fully autonomous robots (see, e.g., [4] for a review) as well as a large variety of artificial-life models of language evolution [5]. A widely used model of autonomous, grounded meaning creation is the discrimination trees model proposed by Steels [6] (see also [7, 8] for applications in language evolution). In this model an individual inhabits a simple world made up of $N$ objects or situations, each of which is described in terms of their features. Feature values are represented by real variables drawn randomly from the uniform distribution in the interval $(0, 1)$. These features are, of course, abstract and have no particular meaning in the model, though it may be helpful to think of them as perceptual features such as color or smell. The individual interacts with the objects by using sensory channels, which are sensitive to the corresponding features of the objects. In particular, there is a specific sensory channel for each feature of the object (e.g., vision for color, olfaction for smell, etc.), which can detect whether a particular value of a feature falls between two bounds.

At the outset, the channels have no discriminatory power - they are sensitive to the entire range of feature values $(0, 1)$. In Steels' model, the individual has the faculty to split the sensitivity range of a channel into two discrete segments, resulting in a discrimination tree. The nodes of this binary tree are then interpreted as categories or meanings. It is the failure to distinguish between any two objects that leads to further splitting or refinement of the discrimination tree and hence to improvement of the semantic structure of the sensory channel. According to Steels [6], this is achieved through repeated discrimination games, in which the individual attempts to distinguish a certain object from a context formed by a random subset of the remaining $N-1$ objects. Whenever a failure occurs a sensory channel is chosen at random, and a randomly-chosen node of its corresponding discrimination tree is split into two new nodes, each one sensitive to half of the range of values of the parent node. Note that the new categories created in this manner may or may not be useful in the discrimination of the objects, since the refinement strategy is completely random. This randomness is an important feature of the model - when the individual is unable to distinguish a particular object from any of the objects that form the context, it has not clue about the feature values of that object, and so it should show no preference for refining any particular sensory channel. After very many such refinements one would expect that, eventually, the individual will develop successful discrimination trees.

Despite the popularity and wide use of Steels' model in robotic applications, even very basic issues, such as the dependence of the expected number of refinements necessary to categorize $N$ objects on the number $M$ of sensory channels, remain unexplored. In fact, as we will show below in the case of a single channel, perfect categorization is unachievable, in a statistical sense, for a finite number of refinements.

In what follows we will consider a variant of the categorization mechanism described above. The main changes are as follows (see Fig. 1). First, we will choose the context of a discrimination game to be the entire set of objects. This allows us to display the values of a given feature of all objects in a line of unit length. There is a line for each feature or sensory channel. Second, at each trial of the discrimination game the individual attempts
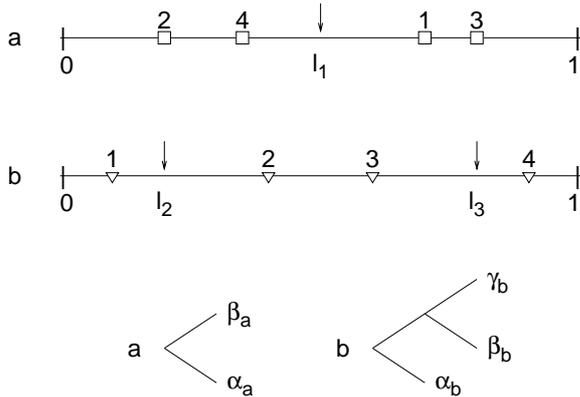
FIG. 1: Illustration of a successful discrimination game for two sensory channels, $a$ and $b$, and $N = 4$ for objects. The values of the object features $a$ and $b$ are represented by the symbols □ and ▽, respectively, and labeled by the object indices. The arrows indicate the discriminatory power of the sensory channels, that can also be represented by discrimination trees. For example, the leaf $\alpha_a$ is sensitive to values of feature $a$ in the range $(0, l_1)$, whereas leaf $\beta_b$ to values of feature $b$ in the range $(l_2, l_3)$.

to categorize all $N$ objects. If it succeeds then the game ends, otherwise one of the sensory channels is refined. Hence the number of trials of the discrimination game, which we will denote by $m$, equals the total number of refinements. Third, the random refinement strategy at trial $m$ of the discrimination game consist of two steps: first we choose a channel at random and then we generate a random number $l_m \in (0, 1)$ that will refine the discriminatory power of the selected channel, as shown in Fig. 1. At the end of the game the whole process can be represented by discrimination trees (one tree for each channel), the leaves of which are sensitive to feature values determined by the ordered set of the random numbers $l_k$ associated to a channel. The final discrimination capability of the tree is determined by its leaves. These changes, while not affecting the essence of the original proposal, allow us to derive analytical results for $N = 2$, and to carry out Monte Carlo simulations for relatively large values of $N$ and $M$.

First let us consider the simplest possible situation: two objects ($N = 2$) and a single channel ($M = 1$). The objects are characterized by the feature values $x_i$, $i = 1, 2$ which are chosen independently from the uniform distribution in the unit interval. In this case, the relevant quantity for the discrimination game is the distance $y = |x_2 - x_1|$, the distribution of which is simply $p(y) = 2(1 - y)$ for $y \in [0, 1]$. As already said, the discrimination game ends when a uniformly distributed random number $l$ is generated such that $l < y$. Given the distance $y$, the probability that this event happens at the $m$th trial is given by the geometric distribution

$(1 - y)^{m-1} y$, with $m = 1, 2, \ldots$, the mean of which is given by the inverse of the probability of a success, $1/y$. Hence the probability that the game halts at the $m$th step regardless of the value of $y$ is

$$Q_m = \int_0^1 dy\, p(y)\, (1 - y)^{m-1}\, y = \frac{2}{(m+1)\,(m+2)}. \quad (1)$$

Introducing the notation $\langle m \rangle_{N,M}$ for the average number of refinements in the case of $N$ objects and $M$ sensory channels we have

$$\langle m \rangle_{2,1} = \sum_{m=1}^{\infty} m Q_m = 2\left(\sum_{m=1}^{\infty} \frac{1}{m} - 1\right) \quad (2)$$

which clearly diverges. Hence, a single sensory channel is insufficient to guarantee discrimination of two (or more) objects. In early simulations, this divergent behavior was mistakenly interpreted as an exponential increase of $\langle m \rangle_{N,1}$ with increasing $N$ [9]. Next we will show how the introduction of more channels remedies this situation.

Assume there are $M$ sensory channels but still two objects, and that their feature values in channel $a$, $x_1^a$ and $x_2^a$, are chosen independently from the uniform distribution, as before. Note that the feature values are statistically independent random variables, regardless of whether they belong to the same or to distinct sensory channels. Hence for each channel we can define the distance $y^a = |x_2^a - x_1^a|$, which is given by the same probability as in the single-channel case. Since at each trial, we choose a sensory channel at random (i.e., with equal probability), the probability of a success (and hence of the end of the game) is $\sum_{a=1}^{M} y^a / M$. Hence,

$$\langle m \rangle_{2,M} = \prod_{a=1}^{M} \int_0^1 dy^a p(y^a) \frac{M}{y^1 + \ldots + y^M} \quad (3)$$

from which we can confirm the (logarithmic) divergence for $M = 1$ and obtain, through the explicit evaluation of the integrals, $\langle m \rangle_{2,2} = 8\,(4 \ln 2 - 1)/3 \approx 4.7269$ in the case of two channels. In general, we can rewrite (3) as

$$\begin{aligned}
\langle m \rangle_{2,M} &= M \int_0^\infty d\xi \left[ 2 \int_0^1 dy\, (1 - y)\, \mathrm{e}^{-\xi y} \right]^M \\
&= M \int_0^\infty d\xi \left\{ \frac{2}{\xi} \left[ 1 - \frac{1}{\xi}\left(1 - \mathrm{e}^{-\xi}\right) \right] \right\}^M. \quad (4)
\end{aligned}$$

In the limit of very many sensory channels ($M \gg 1$) only terms $\xi \sim 1/M$ contribute to the integral yielding thus

$$\langle m \rangle_{2,M} = 3\left(1 + \frac{1}{2M} + \frac{3}{20M^2} + \ldots\right). \quad (5)$$

The case of more than two objects ($N > 2$) is much more complicated. An analytical approach in the line of that presented before seems impossible because now the rules of the discrimination game cannot be described solely
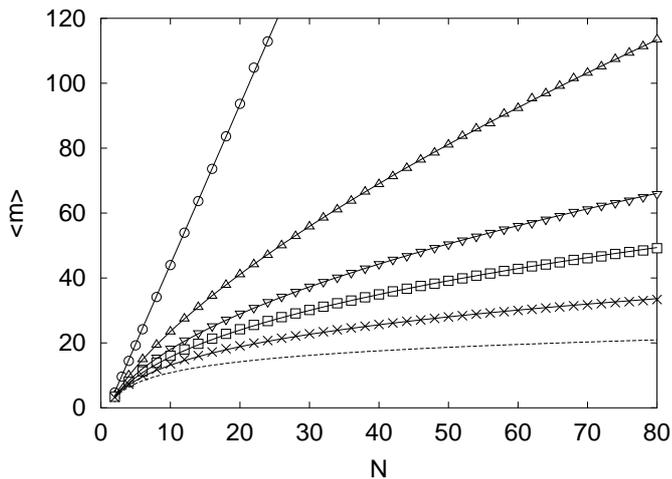
FIG. 2: Average number of trials for perfect discrimination of $N$ objects for $M = 2(\bigcirc), 3(\triangle), 4(\triangledown), 5(\square)$, and $8(\times)$ sensory channels. The solid lines are the numerical fitting (6) and the dashed line is the lower bound obtained in the limit $M \to \infty$.



FIG. 3: Rescaled average number of trials for perfect discrimination $\Lambda$ [see Eq. (7)] as function of $\ln N$. The straight line is the function $\Lambda = \ln N$ and the symbol conventions are the same as in the previous figure.

in terms of the distances between the object features (in which case we could use results of the random ordered intervals [10]): the relative position of each object feature value in a given channel plays a role too. For instance, consider the example illustrated in Fig. 1, for which the feature values are $x_1^a = 0.7, x_2^a = 0.2, x_3^a = 0.8, x_4^a = 0.35$ in channel $a$ and $x_1^b = 0.1, x_2^b = 0.4, x_3^b = 0.6, x_4^b = 0.9$ in channel $b$. Then two trials only (e.g., $l_1 = 0.5$ at $a$ and $l_2 = 0.5$ at $b$) are sufficient to discriminate between the four objects. (Note the minor role played by the distances between feature values in this example.) Therefore, we resort to extensive Monte Carlo simulations of the discrimination games for general $N$ and $M$ in which the results are averaged over $10^7$ independent realizations of the object features. This seemingly exaggerated amount of samples, which makes the sizes of the error bars negligible in comparison to the sizes of the symbols used in the figures, is necessary to obtain reliable estimates of the expected number of refinements for large $N$ and $M$.

The average number of trials of the discrimination game till success $\langle m \rangle$ when the number of channels is fixed and the number of objects is increased is illustrated in Fig. 2. (Henceforth we will use the simpler notation $\langle m \rangle$ in place of $\langle m \rangle_{N,M}$, except when we want to stress that the analysis is valid only for particular values of $M$ or $N$.) An important feature of these results is the slow increase of $\langle m \rangle$ with increasing $N$, which attests the efficiency of the categorization mechanism. More pointedly, the data of Fig. 2 can be fitted by the function

$$\langle m \rangle_{\text{fitting}} = a_M \left( N^{2/M} - 1 \right) \quad (6)$$

with $a_M \approx 2.02M + 0.54$. A better appreciation of the goodness of this fitting is obtained by rescaling $\langle m \rangle$ as

$$\Lambda = \frac{M}{2} \ln \left( 1 + \frac{\langle m \rangle}{a_M} \right) \quad (7)$$
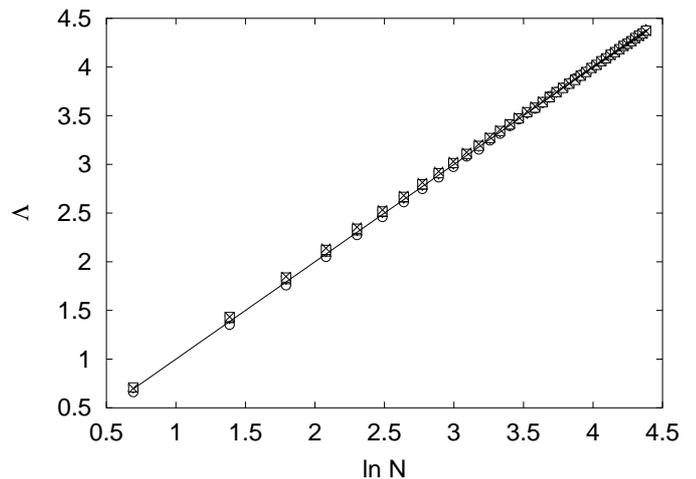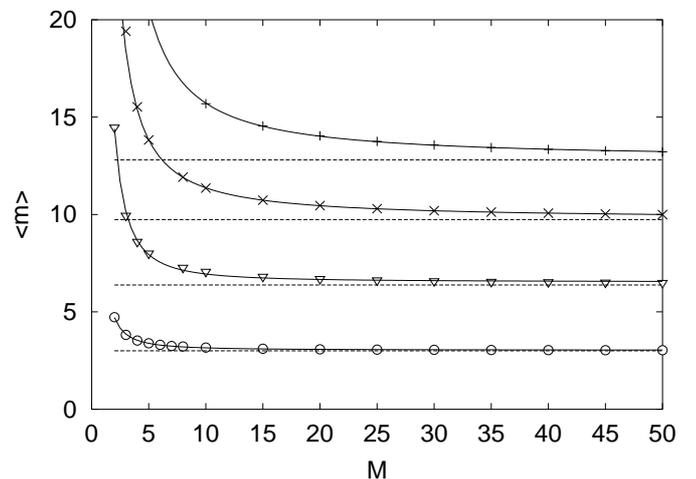


FIG. 4: Average number of trials for perfect discrimination in the case of $M$ channels and $N = 2(\bigcirc), 4(\triangledown), 8(\times)$, and $15(+)$ objects. The solid lines are the quadratic fittings in the variable $1/M$ and the horizontal dashed lines are the estimated asymptotic values that results from those fittings.

and plotting $\Lambda$ against $\ln N$ as shown in Fig. 3. The collapse of the data for different $M$ into a single curve demonstrates that the rescaling (7) is effective to eliminate the dependence on $M$ of the function $\Lambda$. In addition, the unit slope of the straight line that fits the collapsed data supports the validity of the scaling $\langle m \rangle \sim N^{2/M}$ for large $N$. As expected, by increasing the number of channels $M$, we can reduce the number of trials needed to discriminate between the objects. As we will see next, however, the existence of a nonzero lower bound for $\langle m \rangle$ limits the gain of using many sensory channels.

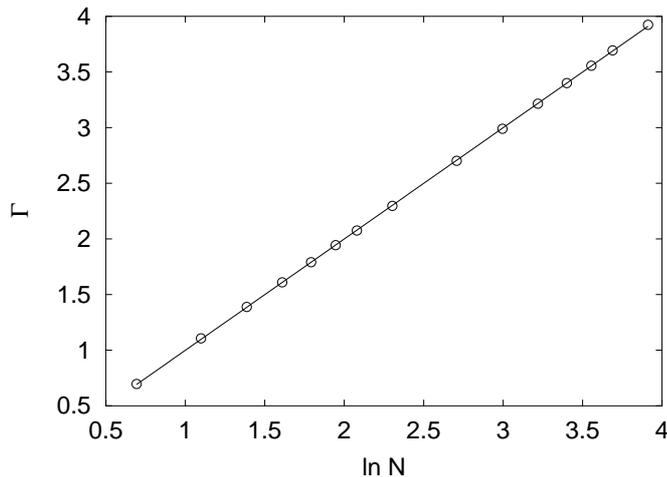To obtain the dependence of $\langle m \rangle$ on $N$ for large $M$

FIG. 5: Rescaled average number of trials for perfect discrimination $\Gamma$ [see Eq. (8)] in the case of infinitely many sensory channels ($M \to \infty$) as function of $\ln N$. The straight line is the function $\Gamma = \ln N$.

(dashed curve in Fig. 2), first we plot $\langle m \rangle$ as function of $M$ for fixed $N$ and then we fit the data using the prescription $\langle m \rangle \approx a_0 + a_1/M + a_2/M^2$, with $a_i = a_i(N), i = 0, 1, 2$, as illustrated in Fig. 4. The choice of this fitting is motivated by the exact solution for the case $N = 2$ given by Eq. (5). The quantity of interest here is the asymptotic value of the number of trials till success $\langle m \rangle_{N,\infty} = a_0(N)$. As could be hinted from

Eqs. (6) and (7), we find that $\langle m \rangle_{N,\infty}$ increases with $N$ as $\ln N$. This can be proved by introducing the function

$$\Gamma = [\langle m \rangle_{N,\infty} + 0.41]/4.89 \qquad (8)$$

and plotting it against $\ln N$, as shown in Fig. 5.

To conclude, we have shown that Steels' perceptually grounded meaning creation mechanism [4, 6–8], which is based on discrimination games to categorize $N$ objects, can be very efficient, provided that the number of sensory channels $M$ is larger than one. In particular, for fixed $M$ and large $N$ we find that the average number of trials of the discrimination game till perfect discrimination, $\langle m \rangle$, increases with $N$ as a power $N^{2/M}$ (see Fig. 2). Since $2/M \leq 1$, the running time of this categorization mechanism increases slower than linearly with the number of objects. For infinitely many sensory channels, we find $\langle m \rangle \sim \ln N$. On the other hand, for fixed $N$ and large $M$ we find that $\langle m \rangle$ decreases with $1/M$ towards a nonzero constant value (see Fig. 4). This limiting value, on its turn, increases logarithmically with increasing $N$.

[1] P. Bloom, *How children learn the meaning of words*, (MIT Press, Cambridge, 2000)
[2] S. Harnard, Physica D **42**, 335 (1990).
[3] F. de Saussure, *Course in General Linguistics.* Translated by Wade Baskin. (McGraw-Hill Book Company, New York, 1966)
[4] L. Steels and J-C. Baillie, Rob. Aut. Syst. **43**, 163 (2003).
[5] A. Cangelosi and D. Parisi, in *Simulating the Evolution of Language* edited by A. Cangelosi and D. Parisi (Springer, London, 2002), pp. 3-28.
[6] L. Steels, in *Proceedings Second International Conference on Multi-Agent Systems* edited by M. Tokoro (AAAI Press, Menlo Park, 1996), pp. 338-344.
[7] A. D. M. Smith, Artificial Life, **9**, 557 (2003).
[8] A. D. M. Smith, in *Proceedings 7th European Conference on Artificial Life* edited by W. Banzhaf, T. Christaller, P. Dittrich, J. T. Kim and J. Ziegler, Lecture Notes in Artificial Intelligence, Vol. 2801 (Springer, Berlin, 2003), pp. 499-506.
[9] J. F. Fontanari and L. I. Perlovsky, in *Proceedings International Conference on Integration of Knowledge Intensive Multi-Agent Systems* edited by C. Thompson and H. Hexmoor (IEEE Press, Boston, 2005), pp. 405-410.
[10] D. E. Barton and F. N. David, J. Roy. Stat. Soc. B **18**, 79 (1956).