# Handling Uncertainty in the Emergence of Social Conventions

Norman Salazar, Juan A. Rodriguez-Aguilar, and Josep Ll. Arcos
*IIIA, Artificial Intelligence Research Institute*
*CSIC, Spanish National Research Council*
*Campus UAB, Bellaterra, Spain*
{*norman,jar,arcos*}*@iiia.csic.es*

## Abstract

*Current computational models for the emergence of conventions assume that there is no uncertainty regarding the information exchanged between agents. However, in more realistic MAS uncertainty exists, e.g. lies, faulty operation, or communication through noisy channels. Hence, within these settings conventions may fail to emerge. In this work we propose the use of self-tuning capabilities to increase the robustness of an emergence mechanism by allowing agents to dynamically self-protect against unreliable information.*

## 1. Introduction

It is commonly agreed that the spreading or propagation of behaviors/knowledge is a main, driving factor in the emergence of conventions. And yet there is a further issue that the literature has not addressed. To the best of our knowledge, it is commonly assumed that there is no uncertainty regarding the information propagated/spread between agents. Hence, each agent assumes that the information spread by other agents is completely reliable. Thus, current models for the emergence of conventions do not consider situations where, for instance, some agents may deliberately lie about the information they propagate, some agents unwillingly propagate wrong information because of bad assessments or misjudgments, or the information propagated contains errors because of noisy communication channels. Therefore, mechanisms for the emergence of conventions must deal with such scenarios.

In this work we argue that propagation-based mechanisms for the emergence of conventions can be extended to help agents agree on conventions despite uncertainty in the information exchanged. Such extension consists in endowing a mechanism with self-tuning capabilities that allow each agent to dynamically adjust its local degree of acceptance of incoming information: the more reliable (and therefore valuable) the information in the past, the more prone to accept it in the future. Through self-tuning agents are granted certain level of self-protection against unreliable incoming information.

Our approach is based on employing and extending the infection-based mechanism (IBM) in [3] as the core convention emergence mechanism. The infection-based mechanism uses its infection operator as a tool to spread behaviors amongst agents. This is a key component in the process of reaching global conventions. Hence, if the information spread through infection is somehow unreliable, reaching a global convention(s) may become impossible. The probability of infection of each agent represents the agent's ability to resist an infection (to resist to incoming information). Thus, the closer the probability to one, the less likely for the agent to be infected. Therefore, by dynamically modifying the infection rate (through the probability of infection, $p_{infection}$) it is possible to deal with uncertainty.

The self-tuning component locally operates in each agent by adjusting its infection rate by some factor related to its performance. This adjustment is based on the dynamic non-uniform mutation operator proposed by Michalewicz in [2]. The adjustment to the probability of infection for each agent is computed as follows:

$$p^{t+1}_{infection} = \begin{cases} p^t_{infection} + \Delta(s, UB - p^t_{infection}) & \text{if X = 0} \\ p^t_{infection} - \Delta(s, p^t_{infection} - LB) & \text{if X = 1} \end{cases} \tag{1}$$

where $\Delta(s, y) = y \cdot (1 - r^{(1-s)^3})$ is a function returning a value within $[0, y]$ such that $\Delta(s, y)$ approaches zero the larger the trend ($s$) of the agent's performance; $X$ is a binary random variable taking any of the two values with equal probability; and $LB$ and $UB$ are the lower and upper bound of the infection probability. Therefore, this update rule has the property of performing smaller adjustments as the trend becomes more steep and bolder adjustments as the trend stagnates.

Through the self-tuning mechanism above each agent tries to locally find an appropriate infection probability. However, with this operator being related to an agent's performance,which is in turn related to the agent's interactions, it is possible that agents in the MAS tune to highly differing infection probabilities. This would have a negative effect, causing an overall MAS instability. We propose that the infection parameter is also the subject of the infection so that agents reach a consensus (agreement) on their probability of infection, hence diminishing the likelihood of instability.

## 2. Empirical Evaluation

Agents in a MAS interact with each other through communications modeled as language games [4]. Each interaction is a communication between an agent playing the role of *speaker*, and another one playing the role of *hearer*, relating to a certain concept $o$. To facilitate communication among agents, each agent has a lexicon, $L_i$, which assigns an external representation (word) to the concepts it needs to employ. Therefore, each agent uses the words in its lexicon to build one-word messages that exchanges with its neighboring agents. The recipient of an agent's message may understand a message or not. This directly depends on the degree of agreement on the lexicons of sender and receiver. To summarize, the mechanics of the game are as follows: (1) agent $s$ selects a concept, $o_s \in O_s$; (2) agent $s$ uses its lexicon, $L_s$, to find the word, $w$, that refers to $o_s$; (3) agent $s$ communicates $w$ to agent $h$; (4) agent $h$ decodes $w$ into a concept $o_h \in O_h$; (5) agent $h$ responds according to its understanding of $o_h$; and (6) the game is successful if $s$ is satisfied by $h$'s response (i.e. if $o_s$ = $o_h$).

The purpose of our experiments is to verify if the infection-based mechanism can be employed so that agents can achieve a so-called perfect communication system [1] under uncertainty. In other words, we pursue lexicons with one-to-one mappings between words and concepts. Each *experiment* consists of 50 discrete event simulations, each one running up to $40000$ time-steps (ticks). Each simulation runs with $1000$ agents over a small-world ( $W_{1000}^{<10>,0.1}$) or scale-free ($S_{1000}^{<10>,-3}$) interaction topology. At the beginning of each simulation, each agent uploads a randomly-generated lexicon. During each simulation, at each time-step agents interact through communications with a randomly selected neighbor. Each agent employs her individual understanding, measured as the number of times she has engaged in a successful communication as a speaker, as her evaluation function. This measure is reset after each incubation period in the infection-based algorithm, namely once the interaction period is over.

To simulate uncertainty, the lexicons exchanged by agents during the infection phase of the IBM were randomly corrupted with probability $p_{corruption} = 0.5$, where corruption consists in changing the word assignments (in the lexicon) of two (different) random objects prior to exchanging it.

Figure 1 shows that no common lexicon emerges when using IBM without self-tuning on a small-world topology. Moreover, the same behavior occurs in the scale-free case. Thus, the IBM is not robust enough to deal with uncertainty. Hence, the need to employ the self-tuning approach proposed in section 1. Figure 2 depicts the results of using IBM along with self-tuning. Observe that in the small-world topology the self-tuning IBM reached a global lexicon convention with one-to-one word mappings (100% specificity) in $\sim$ 16000 ticks. To accomplish this the agents reach an infection probability consensus which endows them with a high infection resistance (a per tick $p_{infection} \simeq 0.98$ average). As for the scale-free topology, the self-tuning also helps it reach a near global convention, but a longer time is needed ($\sim$ 40000 ticks).

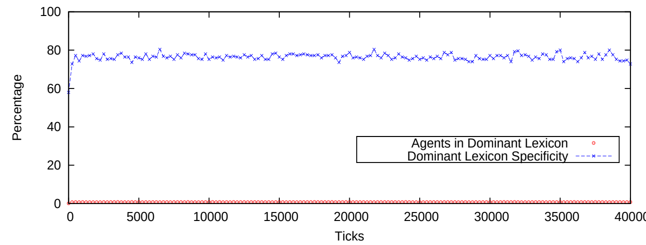


Figure 1. IBM without self-tuning on a small-world.

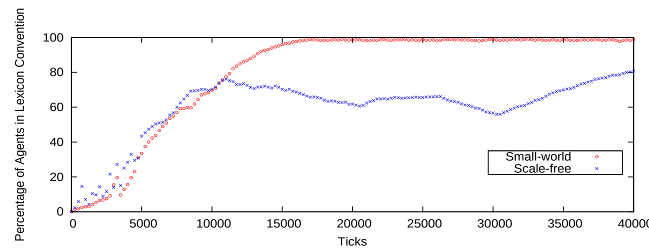


Figure 2. IBM with self-tuning and $p_{corruption} = 0.5$.

## Acknowledgments

N. Salazar thanks CONACyT. JAR thanks JC2008-00337. Funded by TIN2006-15662-C02-01 and CSD2007-0022.

## References

[1] E. D. de Jong and L. Steels. A distributed learning algorithm for communication development. *Complex Systems*, 14:315–334, 2003.

[2] Z. Michalewicz. *Genetic Algorithms+Data Structures =Evolution Programs*. Springer Verlag, 3rd edition, 1996.

[3] N. Salazar, J. A. Rodríguez-Aguilar, and J. L. Arcos. An infection-based mechanism for self-adaptation in multi-agent complex networks. In *In Proceedings of the Second IEEE International Conference on Self-Adaptive and Self-Organizing Systems*, 2008.

[4] L. Steels. Self-organizing vocabularies. In C. G. Langton and K. Shimohara, editors, *Artificial Life V*, pages 179–184, Nara, Japan, 1996.