2009 International Symposium on Nonlinear Theory and its Applications
NOLTA'09, Sapporo, Japan, October 18-21, 2009

# The Origin of Intelligence: A Higher Language Class?

Ruedi Stoop† and Tom Jasa†

†Institute of Neuroinformatics, University of Zürich and ETH Zürich, Irchel Campus, 8057 Zürich
Email: ruedi@ini.phys.ethz.ch

**Abstract**—The question where the computational intelligence associated with humans comes from is an interesting one. General belief states that our supremacy in these matters is the result of our highly-developed language, that is able to define and work with things like concepts and ideas. In the context of formal languages, this question has been formulated in a precise manner by Noam Chomsky. Within that context, a suggestive answer to this problem is that humans possess the most elaborate language structure, which allows the speaker to use constructs like loops and the like. We have tested whether this hypothesis stands the test of the body language of *Drosophila*'s precopulatory courtship behavior and have found it to fail: Seen from several sides, the data appears to share the language classes thought to be exclusively reserved to humans.

## 1. Introduction

A formal grammar $G$ is a quadruple $G = (N, \Sigma, P, S)$, where $N$ is the set of non-terminal symbols, $\Sigma$ the set of terminal symbols, $P$ the set of production rules and $S$ is the start symbol. A rule is applied to a sequence of symbols by replacing an occurrence of the symbols on the left-hand side of the rule with those that appear on the right-hand side. The Chomsky grammar classes -as defined as in Table 1- comprise classes of ever-increasing embracing grammatical complexity $T3 \subset T2 \subset T1 \subset T0$, see Fig. 1.

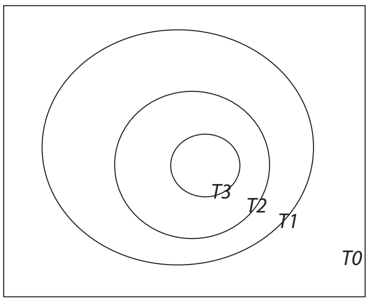


Figure 1: Relation among the Chomsky classes.

Despite the predominance in our perception of spoken language, a great part of the information flow even between humans is not of spoken nature. In this context, the term 'body language' (smell, appearance, gesture, sound) is often used for which, obviously, a corresponding classification may be applied as well. In this body-language, higher grammatical classes would then be more than mere curiosity. Whereas any set of symbols could be used for expression or communication, for efficient communication, this may be too primitive and inexpressive. To be practically useful, grammatical rules, suitable to expressing exact relationships among symbols and to improve signal intelligibility by means of redundancy, are essential. The two properties together allow for the expression from very delicate data relationships up to the encoding of highly nontrivial (e.g., genetic) properties, without running into the latent danger of wrong decoding.

| *grammar: language; automaton; rules* |
|---|
| $T0$ (type-0): recursively enumerable<br>Turing machine<br>$\alpha \to \beta$<br>$\alpha \in V^* N V^*, \ \ \beta \in V^*, \ \ \alpha \neq \varepsilon$ |
| $T1$ (type-1): context-sensitive<br>linearly space-bounded, non-determ.Turing machine<br>$\alpha A \beta \to \alpha \gamma \beta$<br>$A \in N, \ \ \alpha, \beta, \gamma \in V^*, \ \ \gamma \neq \varepsilon$<br>$S \to \varepsilon$ allowed if no rule in P allows $\alpha \to \beta S \gamma$ |
| $T2$ (type-2): context-free<br>non-deterministic pushdown automaton<br>$A \to \gamma, A \to \varepsilon, \quad A \in N, \quad \gamma \in V^*$ |
| $T3$ (type-3): regular<br>finite automaton<br>$S \to \varepsilon$<br>$A \to aB$ (right-regular) or $A \to Ba$ (left-regular)<br>$A \to a, A \to \varepsilon, \quad A, B \in N, \ \ a \in \Sigma$ |

Table 1: Chomsky's language classes.

## 2. Testing for nontrivial grammatical rules

For testing the hypothesis, we investigated data obtained from *D.* pre-copulatory courtship. We used a previously developed framework [1] for measuring and comparing behavior in a fine-grained manner, based on high-speed video clips of the courtship, allow for the resolution of neuronal events. From the physics of complex systems, it is known that the set of closed irreducible orbits provides a complete and unbiased skeleton for the description of the complexity arising from chaotic systems [3, 4, 5, 1]. In this approach, the grammatical rules along which the data are organized are captured in the composition of the various irreducible orbits [3, 4]. For the characterization of behavior, a suitably modified approach needs to be used (see [1], where the validity of the modification is discussed in Ref. [2]). Accordingly, behavior is captured in terms of the set of irreducible closed orbits of fundamental behavioral elements. Biologically, this characterization is corroboratied by the observation that human behavior can be defined as a closed set of well-defined successions of more fundamental actions (in the extreme cases called 'rituals'). We have previously compared [1] the extent by which closed orbits (or the grammar behind it) would be beneficial for the identification of the experimental classes of *D*. behaviors an animal finds itself in. To investigate this, we first characterized the data by means of closed orbits [1] that take account of the grammatical structure in the data. In contrast, in a second model, we characterized each *D*. protagonist by a behavioral vector the entries of which were the natural symbols probabilities (i.e., vectors of length 37). This data model is based on the trivial grammar (no grammatical rules). Remarkably, the characteristic stripes and peaks emerge at the same places most similarly to those obtained by the orbit analysis. In order to work out the similarities/dissimilarities among the classes, closed orbits would therefore not be requested. Both approaches, however, suggest that the experimental class can be determined by a protagonist by a reasonable reliability, but they also suggest that the variability that is nonetheless displayed by the different protagonists is of a non-random origin. The significance of even a small advantage by the grammatical structure, is, however, difficult to assess, as evolutionary, even small enhancements may, over long times, result in an appreciable advantage (see genetic algorithms).

To more thoroughly analyze this issue, we compared our data with surrogate models, based on the given symbol probabilities. In the surrogate data approach, the original data is compared with random models of the underlying symbol probability distribution.The surrogate method will typically provide the most unspecific model compatible with the given distribution, i.e., with the most general, i.e. unrestricted, grammar. Whereas the original closed orbits analysis may be related to a variable-order Markov model approximation of the original data, here the underlying Markov model is of zeroth order. The results obtained for this experiment demonstrate that the basic (dis)similarities between the classes are maintained in this approximation as well. A detailed comparison of the three approaches, confirms that the closed orbits enhance the distinguishability between classes by a very small extent only. The outcome of these experiments implies that closed orbits would not be needed for the recognition of the class membership in the *D*. courtship experiments.

If the closed orbits have a role in efficient communication, by being tuned to describe precise functional relationships in the messages and to improve reading reliability by redundancy, we expect to find a largely increased number of irreducible closed orbits in the original data, indicative of a nontrivial grammar. In a third experiment, we therefore compared how many closed irreducible orbits are present in the original data vs. the surrogate case (in terms of numbers of distinct orbits, not of frequencies). The results displayed in particular in the first two subfigures of Fig. 2, clearly hallmark the presence of a highly developed grammar underlying the courtship language. The investigations also yield that a large portion of the orbits found in the original data are *a priori* unlikely ones. These observations taken together justify the statement that additional, nonaccidental, structure, which can be interpreted as a grammar, underlies *D*. courtship behavior.

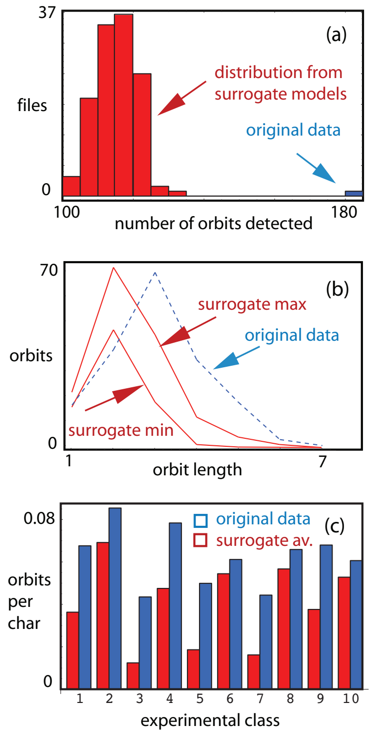


Figure 2: a) Closed orbits in the original (blue) and in 120 surrogate (red) files. b) Maximal/minimal number of orbits of a given length. c) Number of closed orbits per character for the different experimental classes.

## 3. Language class membership of the data

In order to access to what language class *D*. courtship is likely to belong to, we first consider that in our data, we deal with finite information. As a consequence, the language classes $T2 - T0$ is naturally difficult to be discerned. Note, however, that it is generally believed that natural languages fall into the class $T1$ [6]. While thus this question is not properly decidable (the decision problem of class $T1$ is PSPACE-complete), it may, nonetheless, be possible to define some 'generic' construction (class surrogates) patterns that generate distributions (similar to Fig. 2). We may assume that the observed data falls in the class that leads to a minimal distance between the experimental data and the surrogates from a particular class. This approach leads to two questions: What would be a generic generation process? And what would be a suitable distance measure?

In order to address these questions, we started from a random walk model of the data that closely resembles the surrogate data point of view taken above. We first demonstrate that this situation corresponds to a probabilistic type $T3$ grammar, where the term 'probabilistic' means to say that each rule has its own probability of application. The random walk model can then be written as a probabilistic formal language of type $T3$ as follows:

$$G_\omega = (T, V, R, S), \; T = \alpha_\omega, V = \{X\},$$

$$R = \{(R_1, p_1), ..., (R_m, p_m), (X \rightarrow \varepsilon, p_\varepsilon)\},$$

$$i \in \{1, ..., m\}, \; S = X,$$

where $m$ is the number of used rules and where, e.g., $R_1 : (X \rightarrow s_1 X, p_1)$. It is obvious that we deal here with a $T3$-grammar.

We may now use generic strings from the $T3$-grammar and see to what extent an experimental file can be approximated by this model. To this end, we define the distance to the maximum of likelihood of strings, which is the diagonal in the space spanned by the $N$ symbols of the measurement string. If this distance is far out of the bulk of distances generically generated, then we may conclude that the $T3$ generation machinery is not a very likely one. In Fig. 3 we provide two examples of experimental data that can easily modeled / have great difficulty in being modeled by a $T3$-type language. It is useful to capture these differences in

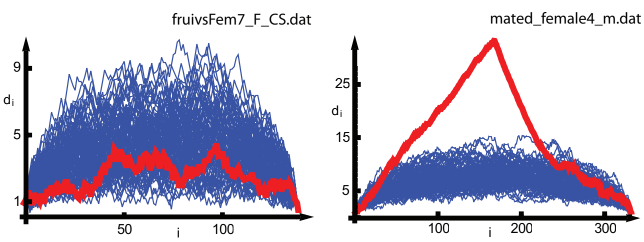


Figure 3: Red lines: distance of experimental, blue lines: distances of 100 surrogate files. Left: $T3$, right: supposedly non-$T3$ grammar.

terms of probabilities rather than distances.

Let $n := \sum_{i \in \{1,2,...,n_{symb},\varepsilon\}} | n_i |$, so that $x = \{n_1, n_2, \ldots, n_{symb}\}$ is the coordinate of a random walk of length $n$ in the symbol space. It is then simple to calculate the probability that a random walk (of length $N$) leads from the beginning over $x$ to some endpoint $x_N$. Using $P_{through}(x) := P_{in}(x) \cdot P_{out}(x)$ and their combinatorial expressions, by plotting $\log(P_{through}(x_i))$ versus $i$ we obtain a better measure of the appropriateness of the $T3$-model with respect to some data. The probability of a string $\omega$ to be generated by the process is then reflected in the logarithm of $H_{through}(\omega) := -\frac{1}{N}\log(P_{through}(\omega)) := -\frac{1}{N}\sum_i \log(P_{through}(x_i))$. Fig. 4 shows the result of this calculation for all files. It is obvious that for some files, the result is compatible with the assumption of a $T3$-grammar, whereas for others, it is not.

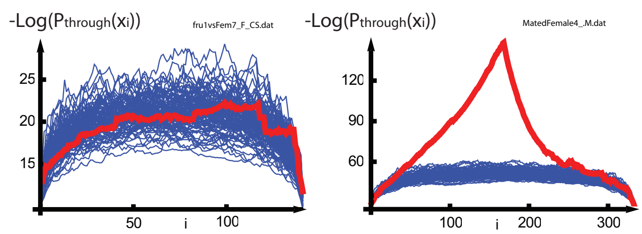


Figure 4: Distance captured in terms of $P_{through}$. Left: $T3$, right: supposedly non-$T3$ grammar

For those that are not, we observe that there is a peak point of largest departure from the given grammar. At this point, we break up the string and model the two parts separately; and likewise iterated, until we come to an end, see Fig. 5. We are able to set up a model that verifies that the grammar type can be inferred from the number $k$ of partial random walks needed to approximate the data:

$$k = 1 \rightarrow T3; \quad k = 2 \rightarrow T2; \quad k \geq 3 \rightarrow T1.$$

## 4. Results and conclusions

The results of the approach are displayed in Fig. 6, where we restrict ourselves to experiments involving genetically 'normal' male animals. It is seen that the female protagonists prefer a $T3$ language, but seem to have the potential of accessing also higher grammar types. Male protagonists generally use a language of higher complexity, with a clear tendency towards $T1$. It emerges that the females have some potential of changing their language as well, somewhat less obvious but similar to what normal males do in the presence of fruitless males. Rather interesting is the possibility indicated by our analysis that the females themselves also convey information. This contradicts a purely decision making role that is exclusively based on the male's performance. Such a view is supported by the

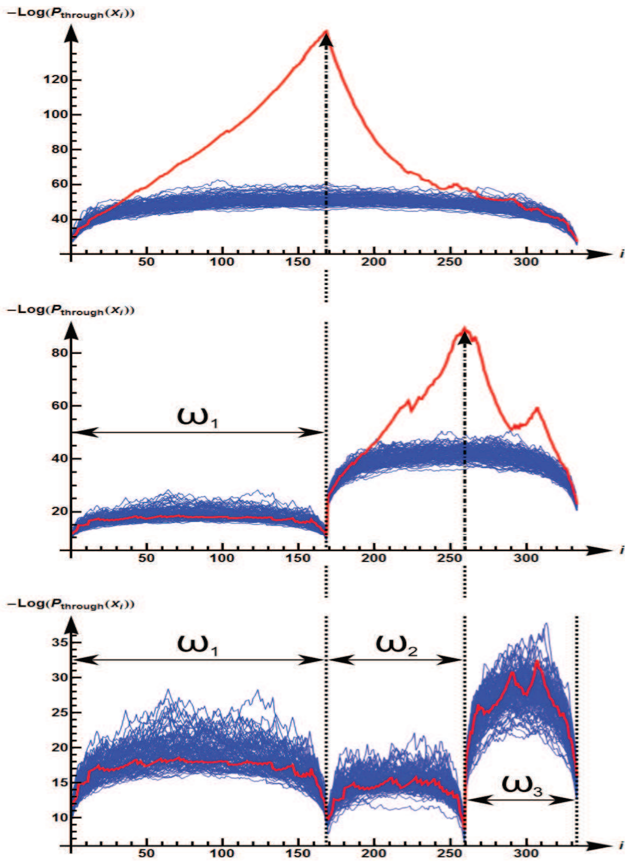


Figure 5: Break-up of a non-$T3$ grammar file.

obtained results in dependence of the protagonist classes. The complexity of the language of *D*. courtship is rather astonishing, apt to express contents of high complexity as would be expected to be needed, e.g., for the transmission of genetic properties. Virgin *D*. appear to test both: communication and decision making. Mature *D*. concentrate on decision making, whereas mated *D*. carefully test for 'better' options, before engaging again. For the male, the situation is a corresponding one: In the presence of virgins, a less elaborate conversation seems to be used. In the presence of mature females, more effort is put into courtship, which is topped only if the female is already mated. Although investigated from a different angle, these results are consistent with what is suggested by the periodic orbit approach and with the hypothesis that during pre-copulatory courtship, detailed information about the protagonists is conveyed (possibly about the genetic properties of a protagonist). Within this context, the presented results may offer a new explanation of the courtship phenomenon among animals.

Within the context of the hypothesis that human intelligence can straight-forwardly be characterized by a language class of increased complexity, our results seem to indicate that such a conclusion would be premature and that still more efforts must be invested in order to clarify one of the most interesting puzzles mankind struggles with: where its computational intelligence comes from.

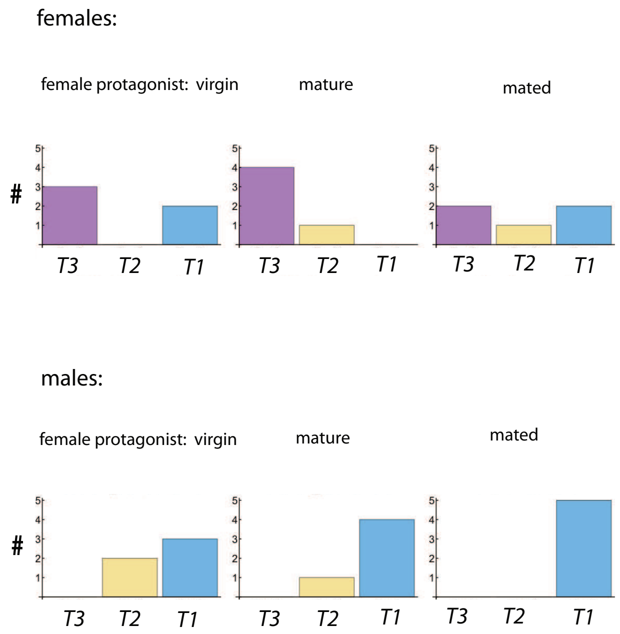


Figure 6: Number of files displaying language complexity classes $T1$, $T2$ and $T3$, for recorded females of state virgin, mature, mated (above), and for recorded males facing females of state virgin, mature, mated (below), indicating unexpected highly complex grammars for both genders.

## References


[1] R. Stoop and B.I. Arthur, *Periodic orbit analysis demonstrates genetic constraints, variability, and switching in Drosophila courtship behavior*, Chaos **18**, 023123; doi:10.1063/1.2918912 (2008).

[2] R. Stoop and B.I. Arthur, *An Expressive Body Language Underlies Drosophila Courtship Behavior*, in Complex Dynamics in Physiological Systems, Springer ISBN 978-1-4020-9142-1 (2009).

[3] P. Cvitanovic, *Periodic orbits as the skeleton of classical and quantum chaos*, Physica D **51** 138 (1991).

[4] P. Cvitanovic, *Dynamical averaging in terms of periodic orbits*, Physica D **83**, 109 (1995).

[5] V.S. Afraimovich and L.A. Bunimovich, *Dynamical Networks: Interplay of Topology, Interactions and Local Dynamics*, Nonlinearity **20**, 1761–1771 (2007).

[6] S.M. Shieber, *Evidence against the context-freeness of natural language* Linguistics and Philosophy **8**, 333–343 (1985).