Eur. Phys. J. B **54**, 127–130 (2006)
DOI: 10.1140/epjb/e2006-00400-x

THE EUROPEAN PHYSICAL JOURNAL B

Statistical analysis of discrimination games

J.F. Fontanari[a]

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

Received 30 June 2006 / Received in final form 27 September 2006
Published online 1st November 2006 –

Abstract. 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 $M \geq 2$.

PACS. 89.75.Fb Structures and organization in complex systems – 02.50.Ey Stochastic processes – 02.50.Le Decision theory and game theory

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 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 discriminating 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 subset of the remaining 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

[a] e-mail: fontanari@ifsc.usp.br

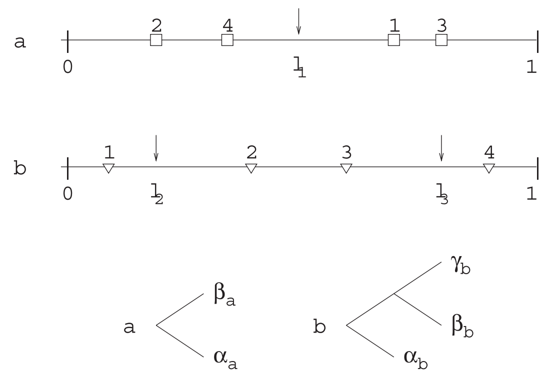

Fig. 1. Illustration of a successful discrimination game for two sensory channels, a and b, and $N = 4$ for objects. The values of the object features a and b are represented by the symbols $\square$ and $\bigtriangledown$, respectively, and labeled by the object indices. The arrows indicate the discriminating power of the sensory channels, that can also be represented by discrimination trees. For example, the leaf α_a is sensitive to values of feature a in the range $(0, l_1)$, whereas leaf β_b to values of feature b in the range (l_2, l_3).

feature of the model for when the individual is unable to distinguish a particular object from the objects in 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 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, denoted by m, equals the total number of refinements. Third, the random refinement strategy at trial m of the discrimination game consists of two steps: first we choose a channel at random and then we generate a uniform random number $l_m \in (0, 1)$ that segments the unit length line into two new (distinct) parts, as shown in Figure 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 in detail 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 $y = |x_2 - x_1|$, since the game ends when a uniformly distributed random number l is generated such that $l < y$. The probability distribution of y, defined by

$$p(y) = \int_0^1 dx_1 \int_0^1 dx_2 \; \delta(y - |x_2 - x_1|), \tag{1}$$

can be readily evaluated and yields $p(y) = 2(1 - y)$ for $y \in [0, 1]$. Given the distance y, the probability that a uniform random number l is smaller than y (i.e., that it falls between x_1 and x_2) is simply y. The probability that this event happens at the mth trial is given by the geometric distribution $(1 - y)^{m-1} y$, with $m = 1, 2, \ldots$ [9]. Hence we can write the probability that the game ends at the mth step regardless of the value of y as

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

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

$$\begin{aligned} \langle m \rangle_{2,1} &= \sum_{m=1}^{\infty} m Q_m = 2 \sum_{m=1}^{\infty} m \left(\frac{1}{m+1} - \frac{1}{m+2} \right) \\ &= 2 \left(\sum_{m=1}^{\infty} \frac{1}{m} - 1 \right) \end{aligned} \tag{3}$$

which diverges due to the presence of the harmonic series. Alternatively, we can evaluate $\langle m \rangle_{2,1}$ by first calculating the expected value of m for a fixed distance y and then carrying out an integration over y, weighting with the probability distribution $p(y)$. Recalling that the mean of the geometric distribution is given by the inverse of the probability of a success ($1/y$, in our case) [9], we can write

$$\langle m \rangle_{2,1} = \int_0^1 dy p(y) \frac{1}{y} = 2 \int_0^1 dy \left(\frac{1}{y} - 1 \right) \tag{4}$$

which also diverges, as expected. Hence, a single sensory channel is insufficient to guarantee discrimination of two (or more) objects. This finding may look counter-intuitive at first sight since one might think that given sufficient trials eventually one will come up with a uniform random number that is smaller than the distance y separating the feature values of the two objects. The problem with this reasoning is that one has to wait $1/y$ trials in the average to generate such a number and since y can be arbitrarily small (note that $p(y)$ reaches its maximum at $y = 0$) the waiting time, when properly averaged over y, becomes infinite. The alternative derivation that resulted in equation (4) makes this point clear. In early simulations, the divergent behavior of $\langle m \rangle_{2,1}$ was mistakenly interpreted as an exponential increase of $\langle m \rangle_{N,1}$ with increasing N [10], which motivated the proposal of an alternative categorization scheme, based on the modeling field theory framework [11], that achieves perfect categorization with a single channel. Next we will show how the introduction of more channels easily remedies this drawback of the discrimination games framework.

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 distributed according to the same probabilitity distribution 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,

$$\langle m \rangle_{2,M} = \prod_{a=1}^{M} \int_0^1 dy^a p(y^a) \frac{M}{y^1 + \ldots + y^M} \tag{5}$$

from which we obtain, through the explicit evaluation of the integrals, $\langle m \rangle_{2,2} = 8\left(4 \ln 2 - 1\right)/3 \approx 4.7269$ in the case of two channels. In general, we can rewrite (5) as

$$\begin{aligned} \langle m \rangle_{2,M} &= M \int_0^\infty d\xi \left[2 \int_0^1 dy\, (1-y)\, \mathrm{e}^{-\xi y} \right]^M \\ &= M \int_0^\infty d\xi \left\{ \frac{2}{\xi} \left[1 - \frac{1}{\xi} \left(1 - \mathrm{e}^{-\xi}\right) \right] \right\}^M . \end{aligned} \tag{6}$$

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

$$\langle m \rangle_{2,M} = 3 \left(1 + \frac{1}{2M} + \frac{3}{20M^2} + \ldots \right). \tag{7}$$

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 in terms of the distances between the object features (in which case we could use the results of the analysis of random ordered intervals [12]): the relative position of each object feature value in a given channel plays a role too. For instance, consider the example illustrated in Figure 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 exagerated 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.

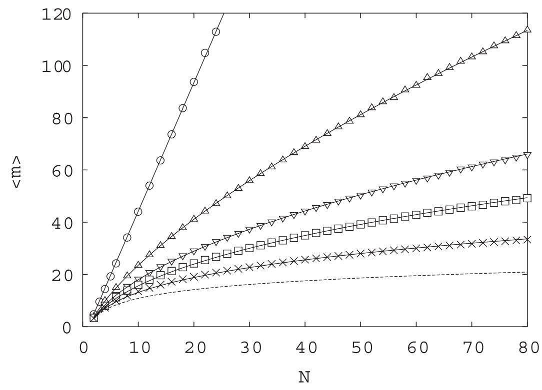

Fig. 2. Average number of trials for perfect discrimination of N objects for $M = 2(\bigcirc), 3(\triangle), 4(\bigtriangledown), 5(\square)$, and $8(\times)$ sensory channels. The solid lines are the numerical fitting (8) and the dashed line is the lower bound obtained in the limit $M \to \infty$.

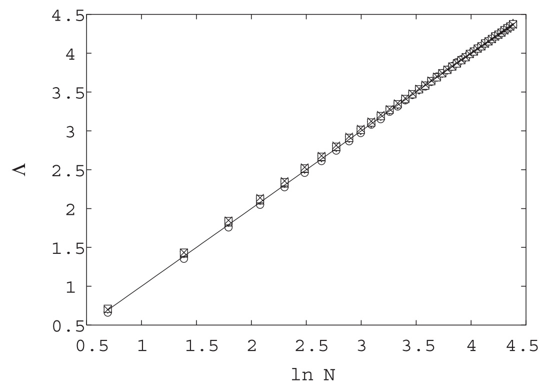

Fig. 3. Rescaled average number of trials for perfect discrimination Λ [see Eq. (9)] as function of $\ln N$. The straight line is the function $\Lambda = \ln N$ and the symbol conventions are the same as in the previous figure.

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 Figure 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 Figure 2 can be fitted by the function

$$\langle m \rangle_{\text{fitting}} = a_M \left(N^{2/M} - 1 \right) \tag{8}$$

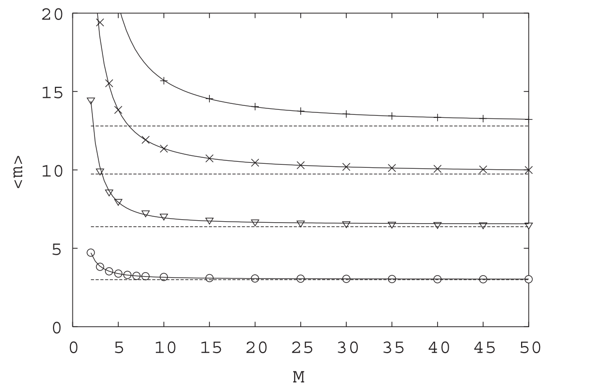

Fig. 4. Average number of trials for perfect discrimination in the case of M channels and $N = 2(\bigcirc), 4(\bigtriangledown), 8(\times)$, and $15(+)$ objects. The solid lines are the quadratic fittings in the variable $1/M$ and the horizontal dashed lines are the estimated asymptotic values that results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

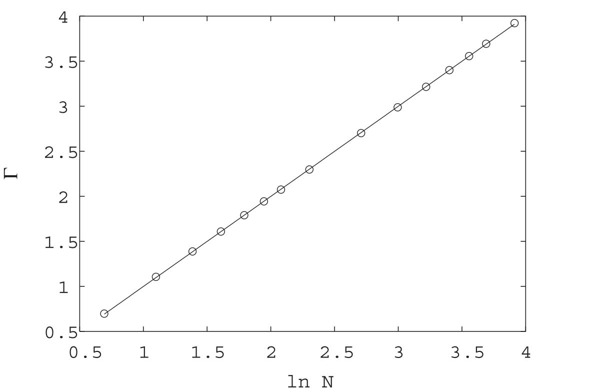

Fig. 5. Rescaled average number of trials for perfect discrimination Γ [see Eq. (10)] in the case of infinitely many sensory channels ($M \to \infty$) as function of $\ln N$. The straight line is the function $\Gamma = \ln N$.

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

$$\Lambda = \frac{M}{2} \ln \left(1 + \frac{\langle m \rangle}{a_M} \right) \tag{9}$$

and plotting Λ against $\ln N$ as shown in Figure 3. The collapse of the data for different M into a single curve demonstrates that the rescaling (9) is effective to eliminate the dependence on M of the function Λ. 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. However, as we will see next, 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 (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 Figure 4. The choice of this fitting is motivated by the exact solution for the case $N = 2$ given by equation (7). 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 equations (8) and (9), we find that $\langle m \rangle_{N,\infty}$ increases with N as $\ln N$. This can be proved by introducing the function

$$\Gamma = \left[\langle m \rangle_{N,\infty} + 0.41 \right] / 4.89 \tag{10}$$

and plotting it against $\ln N$, as shown in Figure 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 sublinearly 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.

Effort sponsored by the Air Force Office of Scientific Research, Air Force Material Command, USAF, under grant number FA8655-05-1-3031. The US Government is authorized to reproduce and distribute reprints for Governmental purpose notwithstanding any copyright notation thereon. This research was supported also by CNPq and FAPESP, Project No. 04/06156-3.

References

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, J.-C. Baillie, Rob. Aut. Syst. **43**, 163 (2003)
5. A. Cangelosi, D. Parisi, in *Simulating the Evolution of Language*, edited by A. Cangelosi, 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, J. Ziegler, Lecture Notes in Artificial Intelligence (Springer, Berlin, 2003), Vol. 2801, pp. 499–506
9. W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. I (Wiley, New York, NY, 1968)
10. J.F. Fontanari, L.I. Perlovsky, in *Proceedings International Conference on Integration of Knowledge Intensive Multi-Agent Systems*, edited by C. Thompson, H. Hexmoor (IEEE Press, Boston, 2005), pp. 405–410
11. L.I. Perlovsky, *Neural Networks and Intellect: Using Model-Based Concepts* (Oxford University Press, Oxford, 2001)
12. D.E. Barton, F.N. David, J. Roy. Stat. Soc. B **18**, 79 (1956)