Cultural route to the emergence of linguistic categories

Andrea Puglisi*, Andrea Baronchelli†, and Vittorio Loreto*‡§

*Istituto Nazionale per la Fisica della Materia–Consiglio Nazionale delle Ricerche (SMC) and Dipartimento di Fisica, ''Sapienza'' Università di Roma, Piazzale A. Moro 2, 00185 Roma, Italy; †Departament de Física i Enginyeria Nuclear, Universitat Politècnica de Catalunya, Campus Nord, Mòdul B4, 08034 Barcelona, Spain; and ‡Fondazione ISI, Viale S. Severo 65, 10133 Torino, Italy

Communicated by Giorgio Parisi, University of Rome, Rome, Italy, March 14, 2008 (received for review January 19, 2007)

Categories provide a coarse-grained description of the world. A fundamental question is whether categories simply mirror an underlying structure of nature or instead come from the complex interactions of human beings among themselves and with the environment. Here, we address this question by modeling a population of individuals who co-evolve their own system of symbols and meanings by playing elementary language games. The central result is the emergence of a hierarchical category structure made of two distinct levels: a basic layer, responsible for fine discrimination of the environment, and a shared linguistic layer that groups together perceptions to guarantee communicative success. Remarkably, the number of linguistic categories turns out to be finite and small, as observed in natural languages.

language dynamics | physics | natural categorization | complex systems

Categories are fundamental to recognition, differentiation, and understanding of the environment. According to Aristotle, categories are entities characterized by a set of properties that are shared by their members (1). A recent wave in cognitive science, however, has operated a shift in viewpoint from the object of categorization to the categorizing subjects (2, 3): categories are culture-dependent conventions shared by a given group. In this perspective, a crucial question is how they come to be accepted at a global level without any central coordination (4–9). The answer has to be found in communication; that is the ground on which culture exerts its pressure. An established breakthrough in language evolution (4, 10–12) is the appearance of linguistic categories; i.e., a shared repertoire of form-meaning associations in a given environment (2, 3, 5, 13–16). Different individuals may in principle perceive, and even conceptualize, the world in very different ways, but they need to align their linguistic ontologies to understand each other.

In the past there have been many computational and mathematical studies addressing the learning procedures for form-meaning associations (17, 18). From the point of view of methodology, the evolutionary scheme, based on the maximization of some fitness functions, has been extensively applied (19, 20). Recent years, however, have shown that also the orthogonal approach of self-organization can be fruitfully exploited in multiagent models for the emergence of language (6–8). In this context, a community of language users is viewed as a complex dynamical system that has to develop a shared communication system (21, 22). In this debate, a still open problem concerns the emergence of a small number of forms out of a diverging number of meanings. For example, the few "basic color terms," present in natural languages, coarse-grain an almost infinite number of perceivable different colors (23–25).

Following this recent line of research, our work shows that an assembly of individuals with basic communication rules and without any external supervision may evolve an initially empty set of categories, achieving a nontrivial communication system characterized by a few linguistic categories. To probe the hypothesis that cultural exchange is sufficient to this extent, individuals in our model are never replaced [unlike in evolutionary schemes (19, 20)], the only evolution occurring in their internal form-meaning association tables; i.e., their "mind." The individuals play elementary language games (26, 27) the rules of which constitute the only knowledge initially shared by the population. They are also capable of perceiving analogical stimuli and communicating with each others (6, 7).

The Category Game Model

Our model involves a population of N individuals (or players), committed in the categorization of a single analogical perceptual channel, each stimulus being represented as a real-valued number ranging in the interval [0, 1].

Modeling Categories. Here, we identify categorization as a partition of the interval [0, 1) in discrete subintervals, from now onwards denoted as "perceptual categories," or simply "categories." This approach can also be extended to categories with prototypes and fuzzy boundaries, for instance adding a weight structure upon it. Typical proposals in the literature, such as prototypes with a weight function equal to the inverse of the distance from the prototype (7), are exactly equivalent to our "rigid boundaries" categories. Moreover, all of the results of our experiment can be easily generalized to multidimensional perceptual channels, provided an appropriate definition of category domains is given. It should be kept in mind that the goal of our work is to investigate why the continuum of perceivable meanings in the world is organized, in language, in a finite and small number of subsets with different names, with a no immediate (objective) cause for a given partition with respect to other infinite possibilities. Apart from the evident example of the partition of the continuous light spectrum in a small number of "basic color terms," this phenomenon is widespread in language: one can ask, for example, what objective differences allow one to distinguish a cup from a glass; one can present a multidimensional continuum of objects able to "contain a liquid" (including also objects given as a prize), but a natural discontinuity between cups and glasses does not appear; our model, even reducing the phenomenon to the case of a one-dimensional continuum, unveils a mechanism that can be easily extended to any kind of space, once it has been provided with a topology. The mechanism we propose for the discrete partition in linguistic subsets (categories) does not depend on the exact nature of this topology, which is of course a fundamental, yet different, matter of investigation.

Negotiation Dynamics. Each individual has a dynamical inventory of form-meaning associations linking perceptual categories (meanings) to words (forms), representing their linguistic counterpart. Perceptual categories and words associated to them co-evolve dynamically through a sequence of elementary com-

---

Author contributions: A.P., A.B., and V.L. designed research; A.P. and A.B. performed research; and A.P., A.B., and V.L. wrote the paper.

[§] To whom correspondence should be addressed. E-mail: vittorio.loreto@roma1.infn.it.

This article contains supporting information online at www.pnas.org/cgi/content/full/0802485105/DCSupplemental.

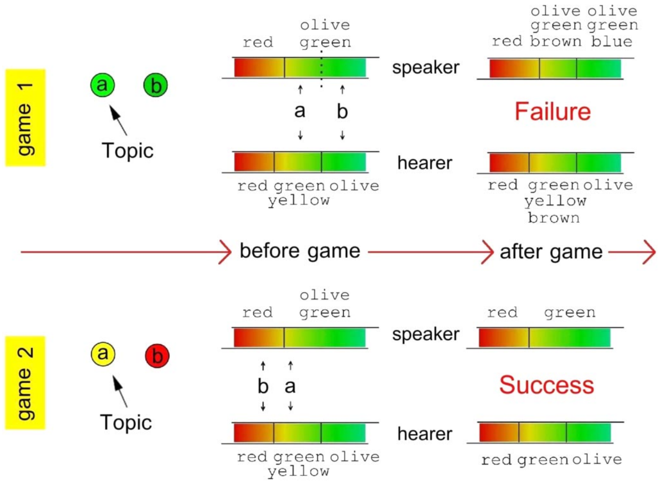

Fig. 1. Rules of the game. A pair of examples representing a failure (game 1) and a success (game 2), respectively. In a game, two players are randomly selected from the population. Two objects are presented to both players. The speaker selects the topic. In game 1, the speaker has to discriminate the chosen topic (''a'' in this case) by creating a new boundary in his rightmost perceptual category at the position (a + b)/2. The two new categories inherit the words inventory of the parent perceptual category (here, the words ''green'' and ''olive'') along with a different brand new word each (''brown'' and ''blue''). Then, the speaker browses the list of words associated to the perceptual category containing the topic. There are two possibilities: if a previous successful communication has occurred with this category, the last winning word is chosen; otherwise, the last created word is selected. In the present example, the speaker chooses the word ''brown'' and transmits it to the hearer. The outcome of the game is a failure because the hearer does not have the word ''brown'' in his inventory. The speaker unveils the topic, in a nonlinguistic way (e.g., pointing at it), and the hearer adds the new word to the word inventory of the corresponding category. In game 2, the speaker chooses the topic ''a,'' finds the topic already discriminated, and verbalizes it by using the word ''green'' (which, for example, may be the winning word in the last successful communication concerning that category). The hearer knows this word and therefore points correctly to the topic. This is a successful game: both the speaker and the hearer eliminate all competing words for the perceptual category containing the topic, leaving ''green'' only. In general, when ambiguities are present (e.g., the hearer finds the verbalized word associated to more than one category containing an object), these are solved making an unbiased random choice.

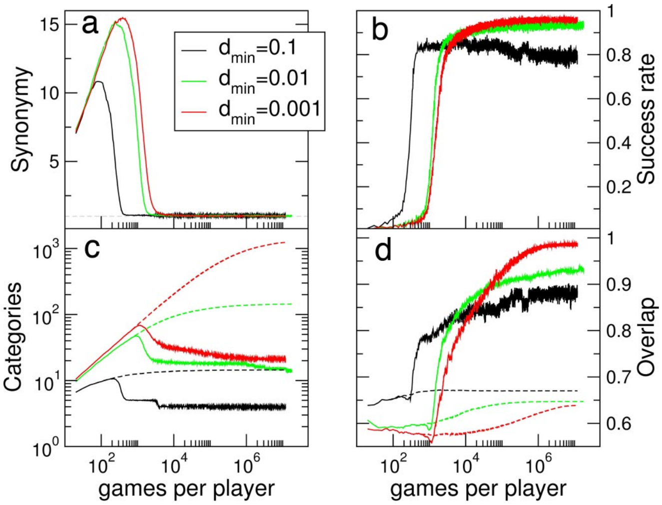

Fig. 2. Results of the simulations with $N = 100$ and different values of $d_{\min}$. (*a*) Synonymy; i.e., average number of words per category. (*b*) Success rate measured as the fraction of successful games in a sliding time windows games long. (*c*) Average number of perceptual (dashed lines) and linguistic (solid lines) categories per individual. (*d*) Averaged overlap—i.e., alignment among players—for perceptual (dashed curves) and linguistic (solid curves) categories.

munication interactions, simply referred to as games. All players are initialized with only the trivial perceptual category [0, 1], with no name associated to it. At each time step a pair of individuals (one playing as speaker and the other as hearer) is selected and presented with a new "scene"; i.e., a set of $M \geq 2$ objects (stimuli), denoted as $o_i \in [0, 1)$ with $i \in [1, M]$. The speaker discriminates the scene, if necessary adding new category boundaries to isolate the topic; then he names one object and the hearer tries to guess it. The word to name the object is chosen by the speaker among those associated to the category containing the object, with a preference for the one that has been successfully used in the most recent game involving that category. A correct guess makes the game successful. Based on the game's outcomes, individuals may update their category boundaries and the inventory of the associated words: in a successful game, both players erase competing words in the category containing the topic, keeping only the word used in that game; in failed games, the speaker points out the topic and the hearer proceeds to discriminate it, if necessary, and then adds the spoken word to its inventory for that category. Detailed example of the game are given in Fig. 1, and the complete algorithmical description is provided in supporting information (SI) *Text*.

The perceptive resolution power of the individuals limits their ability to distinguish objects/stimuli that are too close to each other in the perceptual space: to take this into account, we define a threshold $d_{\min}$, inversely proportional to their resolution power.[¶] In a given scene, the M stimuli are chosen to be at a distance larger than this threshold; i.e., $|o_i - o_j| > d_{\min}$ for every pair (i, j). Nevertheless, objects presented in different games may be closer than $d_{\min}$. The way stimuli are randomly chosen characterizes the kind of simulated environment: simulations will be presented both with a homogeneous environment (uniform distribution in [0, 1]) and more natural environments (e.g., without loss of generality, the distributions of the hue sampled from pictures portraying natural landscapes).

Hierarchical Coordination

The main results of our experiments are presented in Fig. 2. The evolution of the population presents two main stages: (*i*) a phase where players do not understand each other, followed by (*ii*) a phase where communication has reached an averagely high success thanks to the emergence of a common language, still with evolving perceptual categories and a finite fraction of failures due to slightly unaligned categories and ambiguities. The first phase is marked by the growth and decline of synonymy (see Fig. 2*a*). Synonymy, in the context of the "naming game" (a single object to be named), already has been studied (8), and a similar evolution was observed and explained. All individuals, when necessary, create new words with zero probability of repetition: this leads to an initial growth of the vocabulary associated to each perceptual category. New words are spread through the population in later games and, whenever a word is understood by both players, other competing words for the same category are forgotten.[‖] This eventually leads to only one word per category. During the growth of the dictionary, the success rate (see Fig. 2*b*) is very small. The subsequent reduction of the dictionary corresponds to a growing success rate that reaches its maximum

---

[¶] In psychology, $d_{\min}$ is equivalent to the so-called just noticeable difference (JND) or difference limen (DL).

[‖] Extensions of this model can be devised to account for cases where words are not always erased but instead can became more ''specialized,'' eventually yielding to the emergence of a hierarchy of category names.

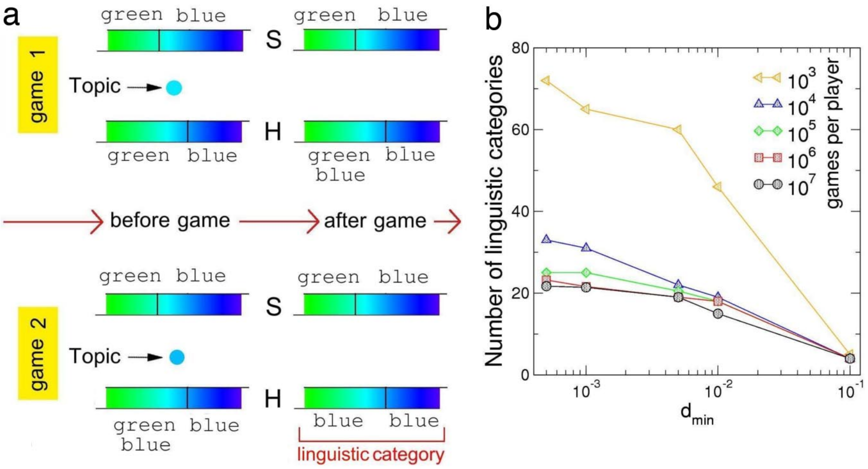

Fig. 3. Saturation in the number of linguistic categories. (*a*) A ''word contagion'' phenomenon occurs whenever the topic falls in a gap between two misaligned categories of two playing individuals. In the shown examples, two individuals play two successive games. In game 1, the speaker (S) says ''blue'' and the hearer (H), unable to understand, adds ''blue'' as a possible word for his leftmost category; successively (game 2), the speaker repeats ''blue'' and the hearer learns this word as the definitive name for that perceptual category; both left and right perceptual categories of the hearer are now identified by the same name ''blue'' and they can be considered (for the purpose of communication) as a single linguistic category. (*b*) Final number of linguistic categories as a function of $d_{\min}$ at different times, with $N = 100$. As the time increases, the number of linguistic categories saturates. At large times, for small $d_{\min}$, the number of linguistic categories becomes independent of $d_{\min}$ itself. Concerning size dependence, only a weak (logarithmic) dependence on N, not shown, is observed.

value after synonymy has disappeared. In all of our numerical experiments the final success rate overcomes 80% and in most of them goes above 90%, weakly increasing with the final number of perceptual categories. Success is reached in a number of games per player of the order of 5×10^2, logarithmically depending on N, and it remains constant hereafter.

The set of perceptual categories of each individual follows a somewhat different evolution (see dashed lines in Fig. 2*c*). The first step of each game is, in fact, the discrimination stage, where the speaker (possibly followed by the hearer) may refine his category inventory to distinguish the topic from the other objects. The growth of the number of perceptual categories n_{perc} of each individual is limited by the resolution power: in a game, two objects cannot appear at a distance smaller than $d_{\min}$ and therefore $n_{\text{perc}} < 2/d_{\min}$. The minimal distance also imposes a minimum number of categories $1/d_{\min}$ that an individual must create before his discrimination process may stop. The average number of perceptual categories per individual, having passed $1/d_{\min}$, grows sublogarithmically, and for many practical purposes it can be considered constant.

The success rate is expected to depend on the alignment of the category inventory among different individuals. The degree of alignment of category boundaries is measured by an overlap function O (defined in *Methods*) that returns a value proportional to the degree of alignment of the two category inventories, reaching its maximum unitary value when they exactly coincide. Its study (see dashed curves in Fig. 2*d*) shows that alignment grows with time and saturates to a value that is, typically, between 60% and 70%; i.e., quite smaller than the communicative success. This observation immediately poses a question: Given such a strong misalignment among individuals, why is communication so effective?

The answer has to be found in the analysis of polysemy; i.e., the existence of two or more perceptual categories identified by the same unique word. Misalignment, in fact, induces a "word contagion" phenomenon. With a small but nonzero probability, two individuals with similar, but not exactly equal, category boundaries may play a game with a topic falling in a misalignment gap, as represented in Fig. 3*a*. In this way, a word is copied to an adjacent perceptual category and, through a second occurrence of a similar event, may become the unique name of that category. Interfering events may occur in-between: it is always possible, in fact, that a game is played with a topic object falling in the bulk of the category, where both players agree on its old name, therefore canceling the contagion. With respect to this canceling probability, some gaps are too small and act as almost perfectly aligned boundaries, drastically reducing the probability of any further contagion. Thus, polysemy needs a two-step process to emerge and a global self-organized agreement to become stable. However, polysemy guarantees communicative success: perceptual categories that are not perfectly aligned tend to have the same name, forming true linguistic categories, much better aligned among different individuals. The topmost curve of Fig. 2*d* displays the overlap function measured considering only boundaries between categories with different names**: it is shown to reach a much higher value, even larger than 90%.

The appearance of linguistic categories is the evidence of a coordination of the population on a higher hierarchical level: a superior linguistic structure on top of the individual-dependent, finer, discrimination layer. The linguistic level emerges as totally self-organized and is the product of the (cultural) negotiation process among the individuals. The average number of linguistic categories per individual, n_{ling} (Fig. 2*c*, solid curves), grows together with n_{cat} during the first stage (where communicative success is still lacking), then decreases and stabilizes to a much lower value. Some configurations of both category layers, at a time such that the success rate has overcome 95%, are presented in Fig. 4, using different sets of external stimuli.

The analysis, summarized in Fig. 3*b*, of the dependence of n_{ling} on $d_{\min}$ for different times makes our findings robust and, to our knowledge, unprecedented. As the resolution power is increased—i.e., as $d_{\min}$ is diminished—the asymptotic number of linguistic categories becomes less and less dependent on $d_{\min}$ itself. Most importantly, even if any state with $n_{\text{ling}} > 1$ is not stable, we have the clear evidence of a saturation with time, in close resemblance with metastability in glassy systems (28, 29). This observation allows

---

**We define the name of a perceptual category as the word that an individual would choose, according to the rules of the model, to communicate about an object discriminated by that category; i.e., the last winning word or the last created word. Of course, if there is a unique word associated with a category (which is most often the case after homonymy has almost disappeared), the definition above identifies that word as the name of the category.

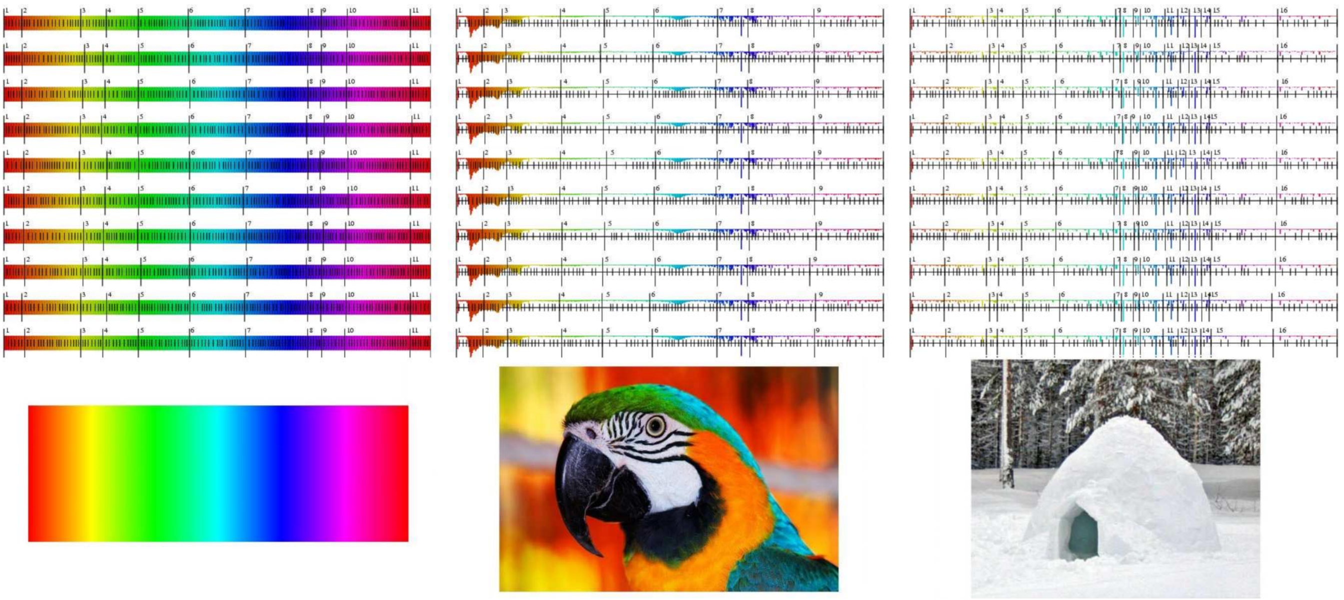

Fig. 4. Categories and the pressure of environment. Inventories of 10 individuals randomly picked up in a population of $N = 100$ players, with $d_{\min} = 0.01$, after 10^7 games. For each player, the configuration of perceptual (small vertical lines) and linguistic (long vertical lines) category boundaries is superimposed to a colored histogram indicating the relative frequency of stimuli. The labels indicate the unique word associated to all perceptual categories forming each linguistic category. Three cases are presented: one with uniformly distributed stimuli (*Left*) and two with stimuli randomly extracted from the hue distribution of natural pictures [*Center* (courtesy of Hamad Darwish) and *Right*]. One can appreciate the perfect agreement of category names and the good alignment of linguistic category boundaries. Moreover, linguistic categories tend to be more refined in regions where stimuli are more frequent: an example of how the environment may influence the categorization process.

one to give a solution to the long-standing problem of explaining the finite (and small) number of linguistic categories n_{ling}. In previous pioneering approaches (6, 7), the number of linguistic categories n_{ling} was trivially constrained (with a small range of variability) by $d_{\min}$, with a relation of the kind $n_{\text{ling}} \propto 1/d_{\min}$, implying a divergence of n_{ling} with the resolution power. In our model, we have a clear indication of a finite n_{ling} even in the continuum limit—i.e., $d_{\min} \to 0$—corresponding to an infinite resolution power.

Conclusions

With the help of an extensive and systematic series of simulations we have shown that a simple negotiation scheme, based on memory and feedback, is sufficient to guarantee the emergence of a self-organized communication system that is able to discriminate objects in the world, requiring only a small set of words. Individuals alone are endowed with the ability of forming perceptual categories, while cultural interaction among them is responsible for the emergence and alignment of linguistic categories. Our model reproduces a typical feature of natural languages: despite a very high resolution power, the number of linguistic categories is very small. For instance, in many human languages, the number of "basic color terms" used to categorize colors usually amounts to ≈10 (23–25), in European languages it fluctuates between 6 and 12, depending on gender, level of education, and social class, while the light spectrum resolution power of our eyes is evidently much higher. Note that in our simulations we observe a reduction, with time, of the number of linguistic categories toward the final plateau. The experimental evidence (30), collected in empirical studies on color categorization, of a growth of the number of categories from technologically less developed societies to more developed ones could be, in our opinion, an effect of the increased number N of players actively involved in the evolution of the communicative process. A plot of n_{ling} versus the number of players N is shown in Fig. S1, to show the effects of finite size on the final category configuration. Finally, we believe that these results could be important both from the point of view of language evolution theories, possibly leading to a quantitative comparison with real data (31, 32) and suggesting new experiments (e.g., different populations sizes and ages), and from the point view of applications [e.g., emergence of new communication systems in biological, social, and technological contexts (33, 34)].

Methods

The degree of alignment of category boundaries is measured by the following "overlap" function:

$$O = 2\sum_{i<j} \frac{O_{ij}}{N(N-1)} \text{ with } O_{ij} = \frac{2\sum_{c_i^j}(l_{c_i^j})^2}{\sum_{c_i}(l_{c_i})^2 + \sum_{c_j}(l_{c_j})^2}, \qquad [1]$$

where l_c is the width of category c, c_i is one of the categories of the *i*th player, and c_i^j is the generic category of the "intersection" set obtained considering all of the boundaries of both players *i* and *j*. The function returns an o_{ij} value proportional to the degree of alignment of the two category inventories, reaching its maximum unitary value when they exactly coincide. A figure making operative this construction is provided as Fig. S2.

ACKNOWLEDGMENTS. We thank G. Andrighetto, A. Baldassarri, T. Belpaeme, C. Cattuto, J. De Beule, E. Polizzi di Sorrentino, L. Steels, and B. De Vylder for many interesting discussions and suggestions, and two anonymous referees for very constructive remarks after a careful reading of the manuscript. This work was partly supported by the European Union under RD Contract IST-1940 (ECAgents) and Contract FP6-IST5-34721 (TAGora). A. Baronchelli acknowledges support from the Departement d'Universitats, Recerca, i Societat de la Informació Generalitat de Catalunya (Spain) and from the Spanish Ministerio de Educación y Ciencia (Fondo Europeo de Desarrollo Regional) through Project FIS2007-66485-C02-01.

1. Barnes J, ed (1995) *The Complete Works of Aristotle: The Revised Oxford Translation*, Bollingen Series (Princeton Univ Press, Princeton), Vol 71, No 2.
2. Lakoff G (1987) *Women, Fire and Dangerous Things* (Chicago Univ Press, Chicago).
3. Gardner H (1987) *The Mind's New Science: A History of the Cognitive Revolution* (Basic Books, New York).
4. Christiansen M, Kirby S, eds (2005) *Language Evolution* (Oxford Univ Press, Oxford).

5. Steels L (1998) The origin of linguistic categories. *The Evolution of Language, Selected papers from the 2nd International Conference on the Evolution of Language*, London.
6. Steels L, Belpaeme T (2005) Coordinating perceptually grounded categories through language: A case study for colour. *Behav Brain Sci* 28:469-529.
7. Belpaeme T, Bleys J (2005) Explaining universal color categories through a constrained acquisition process. *Adapt Behav* 13:293-310.
8. Baronchelli A, Felici M, Caglioti E, Loreto V, Steels L (2006) Sharp transition towards shared vocabularies in multi-agent systems. *J Stat Mech*, P06014.
9. Komarova NL, Jameson KA, Narens N (2007) Evolutionary models of color categorization based on discrimination. *J Math Psychol* 51:359-382.
10. Hurford JR, Studdert-Kennedy M, Knight C, eds (1998) *Approaches to the Evolution of Language: Social and Cognitive Bases* (Cambridge Univ Press, Cambridge, UK).
11. Nowak MA, Komarova NL, Niyogi P (2002) Computational and evolutionary aspects of language. *Nature* 417:611-617.
12. Maynard-Smith J, Szathmary E (1997) *The Major Transitions in Evolution* (Oxford Univ Press, New York).
13. Taylor JR (1995) *Linguistic Categorization: Prototypes in Linguistic Theory* (Oxford Univ Press, Oxford).
14. Coehn H, Lefebvre C, eds (2005) *Handbook of Categorization in Cognitive Science* (Elsevier, New York).
15. Garrod G, Anderson A (1987) Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition* 27:181-218.
16. Labov W (1973) The boundaries of words and their meanings. *New Ways of Analyzing Variation in English*, eds Bailey C-JN, Shuy R (Georgetown Univ Press, Washington, DC), pp 340-373.
17. Hurford J (1989) Biological evolution of the Saussurean sign as a component of the language acquisition device in linguistic evolution. *Lingua* 77:187-222.
18. Oliphant M (2002) Learned systems of arbitrary reference: The foundation of human linguistic uniqueness. *Linguistic Evolution Through Language Acquisition: Formal and Computational Models*, ed Briscoe T (Cambridge Univ Press, Cambridge, UK), pp 23-52.
19. Nowak MA, Krakauer DC (1999) The evolution of language. *Proc Natl Acad Sci USA* 96:8028-8033.
20. Nowak MA, Plotkin JB, Krakauer DC (1999) The evolutionary language game. *J Theor Biol* 200:147-162.
21. Steels L (2000) Language as a complex adaptive system. *Parallel Problem Solving from Nature—PPSN VI*, ed Schoenauer M (Springer, Berlin), pp 17-28.
22. Komarova NL (2006) *Frontiers of Engineering: Reports on Leading-Edge Engineering from the 2005 Symposium* (Natl Acad Press, Washington, DC), pp 89-98.
23. Berlin B, Kay P (1969) *Basic Color Terms: Their Universality and Evolution* (Univ of California Press, Berkeley); reprinted (1991) (Univ of California Press, Berkeley).
24. Lindsey DT, Brown AM (2006) Universality of color names. *Proc Natl Acad Sci USA* 103:16608-16613.
25. Saunders BAC, van Brakel J (1997) Are there non-trivial constraints on colour categorization? *Behav Brain Sci* 20:167-179.
26. Wittgenstein L (1953) *Philosophical Investigations* (Blackwell, Malden, MA).
27. Steels L (1996) Self-organizing vocabularies. *Artificial Life V: Proceedings of the Fifth International Workshop on the Synthesis and Simulation of Living Systems*, eds Langton CG, Shimohara T (MIT Press, Cambridge, MA), pp 179-184.
28. Mezard M, Parisi G, Virasoro MA (1987) *Spin Glass Theory and Beyond* (World Scientific, Teaneck, NJ).
29. Debenedetti PG, Stillinger FH (2001) Supercooled liquids and the glass transition. *Nature* 410:259-267.
30. Kay P, Maffi L (1999) Color appearance and the emergence and evolution of basic color lexicons. *Am Anthropol* 101:743-760.
31. Kay P, Cook RS (2002) *The World Color Survey*, www.icsi.berkeley.edu/wcs.
32. Selten R, Warglien M (2007) The emergence of simple languages in an experimental coordination game. *Proc Natl Acad Sci USA* 104:7361-7366.
33. Steels L (2006) Semiotic dynamics for embodied agents. *IEEE Intell Syst* 21:32-38.
34. Cattuto C, Loreto V, Pietronero L (2007) Semiotic dynamics and collaborative tagging. *Proc Natl Acad Sci USA* 104:1461-1464.