On the origin of the hierarchy of color names

Vittorio Loreto[a,b], **Animesh Mukherjee**[b,c], **and Francesca Tria**[b,1]

[a]Dipartimento di Fisica, Sapienza Universita' di Roma, Piazzale Aldo Moro 5, 00185 Rome, Italy; [b]Institute for Scientific Interchange, Viale Settimio Severo 65, 10133 Turin, Italy; and [c]Department of Computer Science and Engineering, Indian Institute of Technology, Kharagpur 721302, India

Edited by Giorgio Parisi, University of Rome, Rome, Italy, and approved February 16, 2012 (received for review August 14, 2011)

One of the fundamental problems in cognitive science is how humans categorize the visible color spectrum. The empirical evidence of the existence of universal or recurrent patterns in color naming across cultures is paralleled by the observation that color names begin to be used by individual cultures in a relatively fixed order. The origin of this hierarchy is largely unexplained. Here we resort to multiagent simulations, where a population of individuals, subject to a simple perceptual constraint shared by all humans, namely the human Just Noticeable Difference, categorizes and names colors through a purely cultural negotiation in the form of language games. We found that the time needed for a population to reach consensus on a color name depends on the region of the visible color spectrum. If color spectrum regions are ranked according to this criterion, a hierarchy with [red, (magenta)-red], [violet], [green/yellow], [blue], [orange], and [cyan], appearing in this order, is recovered, featuring an excellent quantitative agreement with the empirical observations of the WCS. Our results demonstrate a clear possible route to the emergence of hierarchical color categories, confirming that the theoretical modeling in this area has now attained the required maturity to make significant contributions to the ongoing debates concerning language universals.

color hierarchy | complex systems | computational cognitive science | statistical physics | category game

Color naming represents a paradigmatic problem in cognitive science and linguistics (1–3) due to the unique complex interplay between perception, conceptualization, and language it features. In addition, color naming constitutes an outstanding example of the long "nature versus nurture" debate in cognitive science, namely whether color names are pure arbitrary linguistic conventions (4) (i.e., nurture) or they are coded in some innate human feature (i.e., nature) (5). Color naming patterns exhibit structural regularities across cultures (6–8). Extensive studies involving basic color names have been performed in the past that reveal interesting properties such as the nonrandom distribution of color terms (9) and an optimal partition of the color space by these terms (10). The data gathered in the World Color Survey (WCS) (11), extending the pioneering work by Berlin and Kay (6), provided evidence for the existence of universals in color categorization. Since then, a long line of research (9, 12–16) confirmed the existence of such universals, although the scientific debate is still wide open (17, 18). Recently (19), it has been pointed out how the observed recurrent patterns in language organization could be explained as stable engineering solutions reflecting cultural and historical factors (15) as well as the constraints of human cognition. Along this same line, recent findings (16) suggest how a pure cultural negotiation process, with a slight non-language specific bias, can account for the observed regularities across different populations.

One of the most crucial observations related to the universality of color naming is the existence of basic color names across languages (6). These basic color names are identified (not without ambiguities) as being monolexemic, highly frequent, and agreed upon by speakers of the same language. A surprising experimental finding about color names is the existence of a hierarchy of basic color names which began to be used by individual cultures in a relatively fixed order (6). According to this observation, basic color names can be organized into a coherent hierarchy around the universal focal colors black, white, red, green, yellow, and blue always appearing in this specific order across cultures. The meaning of this implicational hierarchy is as follows: If a population has a name for red, it also has a name for black and white (but not vice versa), if it has a name for green, it also has a name for red (but not vice versa), and so on. It should be remarked that the terms black and white appear in this hierarchy with a meaning close to the general panchromatic English terms dark and light or dull and brilliant rather than equivalent to the specific achromatic terms black and white (we refer to the *SI Text* for a more detailed discussion of the empirical observations). The origin of the observed hierarchy is largely unexplained and the aim of this paper is that of providing a first coherent and quantitative explanation of this phenomenon.

Color categorization has been used as a reference problem in computational studies on symbol grounding where one investigates how a population of interacting individuals can develop a shared repertoire of categories from scratch (12). It has been recently shown how a pure cultural negotiation dynamics, in the form of repeated language games (20–22) called the Category Game (CG) (15), can lead to the coevolution of a shared repertoire of categories and their linguistic labels. The CG considers a simplified representation of the color space consisting in a reduction of the true three-dimensional space to the one-dimensional hue color wheel, neglecting in this way the saturation and brightness dimensions. This abstraction is common in literature (13, 14, 23–25) where often a discrete division of the hue dimension has been adopted to represent the visual space. The novelty introduced by the CG consists in the introduction of a truly continuous perceptual space (e.g., the visible light spectrum) with no predefined category structure. Remarkably, even while the perceptual space is a continuum (as in colors), the emergent number of linguistic labels is finite and small (15), as observed in natural languages. In addition, though the reduction to the hue color wheel seems a very crude assumption, it should be remarked that individuals simulated in the CG, and endowed with the human Just Noticeable Difference (JND) function (27, 28), are able to bootstrap a color categorization whose statistical properties turn out to be in very good quantitative agreement with those observed in the WCS data (16). The JND function describes the variability of the resolution power of human vision with the frequency of incident light (Fig. 1). It is remarkable how a weak non-language specific bias common to all human beings, such as the human JND, can lead to a qualitative and, most strikingly, quantitative agreement with the experimental findings of the WCS. Finally the CG features a dynamical behavior characterized by the persistence of long-lasting metastable states (26). This observation formalizes the intuition that languages change thanks to,

APPLIED PHYSICAL SCIENCES

PSYCHOLOGICAL AND COGNITIVE SCIENCES

Author contributions: V.L., A.M., and F.T. designed research, performed research, contributed new reagents/analytic tools, analyzed data, and wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. E-mail: tria@isi.it.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1113347109/-/DCSupplemental.

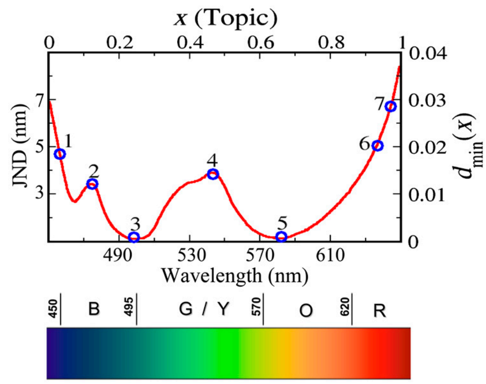

Fig. 1. The JND function. The wavelength change in a monochromatic stimulus needed to elicit a particular JND in the hue space. For the purpose of the CG, $d_{\min}(x)$ and topic, respectively, refers to the JND and the monochromatic stimulus rescaled within the $[0, 1)$ interval. The blue circles represent the centers of seven regions (to be used later in the article) that can be together expressed as a vector c with entries $c_1, c_2, \ldots, c_7$. The specific values for these entries are $c = [0.0301, 0.125, 0.250, 0.465, 0.66015, 0.925, 0.970]$. Each entry, in turn, respectively corresponds to a wavelength (nanometer) that can be written (approximately) as [445, 475, 500, 545, 585, 635, 645].

and notwithstanding, their being the outcome of a collective behavior.

The central result of this paper is that a clear hierarchy for color names is found to naturally emerge, in the framework of the CG (15), through purely cultural negotiations among a population of coevolving agents, each endowed with the human JND function (27, 28). In particular, a hierarchy emerges that ranks different color names with respect to the time needed for them to get fixed in a population. Those names on which the population reaches a faster agreement turn out to be the basic color names, and the order of their emergence reflects the hierarchy found in ref. 6. This finding immediately provides a cultural definition of the basic color names with a degree of explicitness that has not been achieved so far.

The Hierarchical Category Game

The CG (15) (see *Methods*) describes the emergence of a hierarchical category structure made of two distinct levels: a basic layer, responsible for the fine discrimination of the environment (perceptual categories), and a shared linguistic layer that groups together perceptions to guarantee communicative success (linguistic categories). In the version presented here, the CG is further extended to allow for the emergence of a series of shared linguistic layers each of which could guarantee communicative success in progressively more complex tasks.

At each time step, a pair of individuals (one will be denoted as the speaker and the other as the hearer) is randomly selected from the population to play a language game. The speaker has to communicate about a scene composed of, say, two objects. He selects one of these, denoted as the topic, and decides to speak about it. The ensuing language game allows both players to coevolve the structure of their categories as well as their form-meaning inventories. Although the number of perceptual categories is tuned by a parameter of the model that encodes the JND (see *Methods*) and can be arbitrarily large, the number of linguistic categories, in all the emergent layers, turns out to be finite and small, as observed in case of color names for natural languages. A higher and a more refined linguistic layer is accessed by the agents only if, in a game, both the topic and the object have the same name. This homonymy generates a "confusion" in differentiating between the two, possibly resulting in a failure in communication, also referred to as "failure with name" (see *Methods*). Note that access to a higher level requires a high consensus in the adjacent lower level because a significant alignment among the agents is necessary to cause a failure with name. Whereas the first layer of linguistic categories (also reported in refs. 15, 16, and 26) can be associated to the emergence of primary color names, the successive layers might be linked to the emergence of complex color names when the knowledge of the primary color names is not enough to achieve a reasonable communicative success (one can think of a linguistic community comprising specialized individuals, as for instance painters, textile and cosmetic manufacturers; ref. 29).

Dynamics

Let us start our analysis by monitoring the time evolution of the number of linguistic categories at the different levels of the hierarchy. Fig. 2 *A* and *B* report the average number of perceptual and linguistic categories (at different levels) in the population when the model is informed respectively with the average human JND ($\bar{d}_{\min} = 0.0143$) and the actual human JND function [i.e., $d_{\min}(x)$ as shown in Fig. 1]. For the linguistic categories, two regimes are clearly identified in all the different levels. Initially, corresponding to a series of uncorrelated games, the average number of linguistic categories per individual (n_{ling}) exhibits a rapid growth due to the pressure of discrimination (for a detailed description of CG, we refer to *Methods*), followed by a fast drop due to the onset of consensus and the merging of perceptual categories. A second regime eventually emerges in all the levels, characterized by a quasi-arrested dynamics. This slow dynamics is signaled by a "plateau" region, as also observed earlier for level 0 in ref. 26. Interestingly, when the human JND function is adopted, one observes the possibility of the emergence of a third level, although still in its transient phase even after a billion games per player (level 2 in Fig. 2*B*), whereas there is no such signal with a flat JND. This observation implies that, for an average JND, individuals need just two levels to express a category, for the actual JND, individuals need further levels to achieve communicative success. It turns out, in fact, that further levels are typically needed for small values of $d_{\min}$ (i.e., higher resolu-

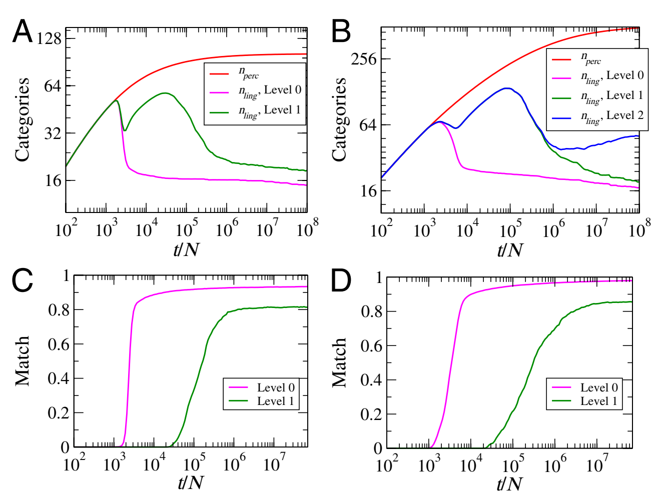

Fig. 2. Emergence of the category structure. (*A*) The average number of perceptual categories as well as the average number of linguistic categories at different levels versus the number of games per player t/N (where t indicates the number of games and N the population size) when JND is set to $\bar{d}_{\min}$. (*B*) The average number of perceptual categories as well as the average number of linguistic categories at different levels versus t/N when JND is set to $d_{\min}(x)$. (*C*) The evolution of the average match in the population at different levels when JND is set to $\bar{d}_{\min}$. The average match (see *Methods* for details) is a measure of the global alignment of the linguistic categories in the whole population. (*D*) The evolution of the average match in the population at different levels when JND is set to $d_{\min}(x)$. Here $N = 500$ and the results represent an average over 30 simulation runs.

tion power) and the actual JND features large regions below its average value (see *SI Text* for further discussion on this issue). In order to quantify the level of alignment of the linguistic categories across the population, we monitor (Fig. 2 *C* and *D*) the emergence of the average pairwise match (see *Methods*) as a function of time (expressed as games per player) for the different levels. Again we repeated the experiment with the average human JND ($\bar{d}_{\min} = 0.0143$) and the actual human JND function [i.e., $d_{\min}(x)$]. Remarkably, the extent of agreement among the agents (as measured by the match) exceeds 95% in level 0 and 80% in level 1. Note that these results are consistent across different population sizes as shown in the *SI Text*.

It is important to remark that the timescales associated with the CG dynamics represent the times of persistence of a particular category in the population. The emergent asymptotic categorization corresponds to a metastable state where global changes are always possible, though progressively less likely as the system ages, which is typically synthesized by saying that the response properties of the system depend on its age (26). This perspective allows us to reconcile the evidence that languages do continuously change still remaining stable enough to be intelligible across a population.

Finally, a possible way to relate the different levels of the category structure to the process of human learning could be as follows. "Level 0" typically refers to the early stages of learning where a linguistic community attempts to agree upon a set of (basic) color terms needed for successful communication. However, as time goes by, the community would naturally feel the need of communicating through more complex color terms (e.g., color of a lipstick or a garment or a car). In the initial stages of this phase the community shall almost surely encounter difficulties in discriminating and communicating about close shades or nuances of color in a scene (analogous to failure with name); however, a second level of agreement could soon emerge within the community, when most of the language speakers are able to resolve and correctly associate higher order color terms to the various objects of the scene and this, in turn, is equivalent to "level 1" in the CG framework.

Hierarchy of Fixation Times

We now focus on the frequency of access to higher levels of linguistic categorization as a function of the local value of the JND. To this end, we report in Fig. 3 a scatter plot of the logarithm of the time (again expressed as games per player) at which the agents need to access level 1 versus the value of the topic in that particular game. The result clearly demonstrates that the agents need to access the higher level early in regions corresponding to high values of $d_{\min}(x)$, whereas they access it quite late in regions corresponding to low values of $d_{\min}(x)$. This observation indicates that an agreement at level 0 is reached faster in regions with high values of $d_{\min}(x)$, resulting in more cases of failure with name in these regions, thereby, forcing the agents to access level 1.

In order to further verify the above observation, we compute the extent of the emergent agreement (i.e., match) at different regions of the perceptual space in level 0. In Fig. 1, the blue circles indicate the centers c_i of seven such regions (i.e., the points of inflection in the JND function) that we choose to calculate the so-called "regional" agreement. We define a region by the length spanning $[c_i - d_{\min}(c_i), c_i + d_{\min}(c_i)]$ where $d_{\min}(c_i)$ is the y value corresponding to the x value c_i (see Fig. 1). In Fig. 4 *A* and *B*, we respectively show, for $N = 500$ and 700, the regional agreement for these seven regions at level 0 (also see the *SI Text*). The plots clearly signal that consensus emerges first in regions corresponding to high values of $d_{\min}$ (e.g., regions 6 and 7) whereas it occurs later in regions corresponding to very low $d_{\min}$ (e.g., regions 3 and 5). Most strikingly, if the regions are arranged according to the time (i.e., t/N) to reach a desired level of consensus (say a match value of 0.1), then they get organized into a hierarchy (Fig. 4 *C* and *D*) with [red, (magenta)-red], [violet], [green/yellow], [blue], [orange], and [cyan] (or [cyan] and [orange] as is usually observed for secondary basic color names) appearing in this order. This result is strikingly similar to that reported in ref. 6. Further, the data points for the fixation times are observed to obey a simple functional form, $Ae^{-\alpha t}$, where A and α are nonzero positive constants (gray lines in Fig. 4 *C* and *D*). In other words, the fixation time for specific primary colors at the population level diverges logarithmically with the resolution power $1/d_{\min}$. Though this specific prediction cannot be checked with the currently available data, it is reminiscent of the logarithm law which is typically associated to human perception. Error bars in Fig. 4 *C* and *D*, representing the intrinsic variability of fixation

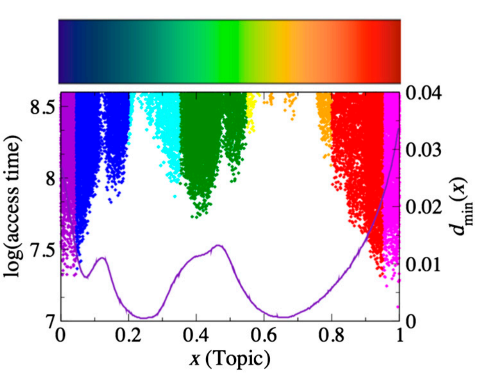

Fig. 3. Activity for different topics. Scatter plot of the logarithm of the time (expressed as games per player) at which the agents need to access level 1 versus the value of the topic in that particular game. The points are colored in a way that they best represent the corresponding region of the visible spectrum. The human JND function [i.e., $d_{\min}(x)$ versus the topic value] and the visible spectrum are given as references. Here $N = 500$ and the results represent an average over 30 simulation runs.

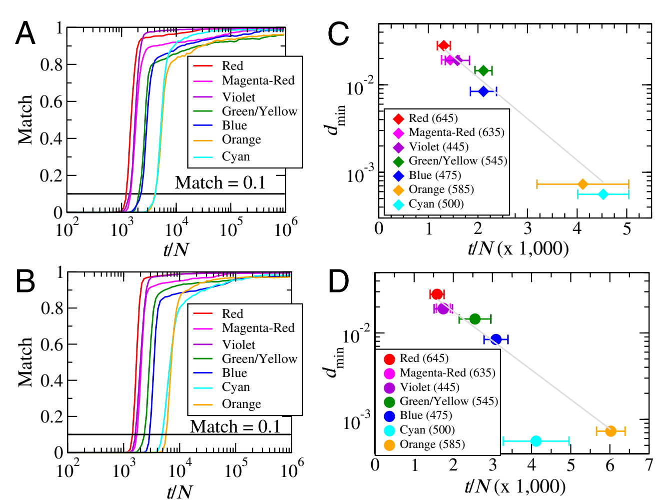

Fig. 4. Agreement emergence. Emergence of the agreement in the population in level 0. Match for (*A*) $N = 500$ and (*B*) $N = 700$ in the seven regions marked in Fig. 1. For better visualization, each curve is plotted in a color that best represents the corresponding region in the hue space (see Fig. 1). The time (i.e., t/N) for (*C*) $N = 500$ and (*D*) $N = 700$ to reach a desired consensus (match = 0.1) versus the value of $d_{\min}$ corresponding to the seven regions. The results present an average over 60 simulation runs. In both the plots, the approximate wavelength (nanometer) associated with each colored data point is mentioned within the parenthesis. Error bars are drawn according to the variance of the distribution of consensus times in the different simulations. The gray lines in both the plots represent a fit of the respective data with an exponential function of the form $Ae^{-\alpha t}$ (see text for more details).

times in different simulations, are important to explain the slight fluctuations in the color name hierarchy as observed in the WCS across different cultures.

It is important to observe how the similarity of the ranking of fixation times obtained in the framework of the CG with that observed in the framework of the WCS is not the outcome of a pure coincidence. It turns out that only a right choice of JND function, coupled with the language game dynamics, can reproduce the color hierarchy observed across human languages. In the *SI Text*, we report the outcomes of two additional experiments performed by substituting the human JND with a flat and an inverse JND. In none of these two cases the hierarchy obtained from the WCS could be reproduced.

Discussion

The two specific exceptions of the observed hierarchy from that suggested in ref. 6 are (*i*) emergence of violet in this hierarchy which is absent in ref. 6 and (*ii*) absence of brown in this hierarchy which appears immediately after blue in ref. 6. This discrepancy can be perhaps explained in the light of the past literature on basic color names. According to Kay and McDaniel (30), both of these color names are secondary basics and can therefore be expressed as fuzzy combinations of the six focal colors. In order to understand the presence of violet in the hierarchy, one needs to concentrate on the second stage of the color lexicon evolution suggested by Berlin and Kay (6) that marks the emergence of red. The authors themselves note that at this point the name red also includes the other end of the spectrum which is primarily violet. In fact, low-wavelength light (perceptually violet), although being at the opposite end of the spectrum, is in many cases perceived as reddish (31) and this is possibly why we see the emergence of violet just after red in the hierarchy (see also the *SI Text* for a further experiment concerning "red" and "violet"). On the other hand, brown is not a spectral color itself and usually refers to a combination of the high-wavelength hues: yellow, orange, or red. Therefore, the term brown can cover a wide range of the visible spectrum mostly inclusive of the different shades of orange and, in particular, is frequently recognized as dark orange (32). Consequently, it may be well argued that the emergence of orange in the hierarchy actually also marks the emergence of brown.

Further, we also note that no evident hierarchy is observed for the linguistic categories at level 1 (see the *SI Text*). Because color names associated to higher linguistic levels are intuitively associated to nonbasic color names, this observation implies that it is hard to arrange complex color names in a clear hierarchy as for the basic color names.

Another important point that deserves mention here is that, although in the current work categorization is invariably associated with naming, nonverbal perceptual learning is equally possible within a population and it has been extensively studied in refs. 33–36. However, our intention here was to seek a suitable answer to a long-standing chicken and egg problem in cognitive science: To name a category, it seems that this category should be already existing and be shared in the population, so how can naming influence the shape of the emerging category structure? The CG is an attempt to show that coordination in a population is possible through a purely structural coupling between the categorization and the naming processes. The emergent patterns allow us to conclude that this coupling is indeed possible and that there is at least some role that a language plays to give rise to the coordination of the perceptually grounded categories. Thus, our contribution here is a plausible solution to the chicken and egg problem through the introduction of a complex interplay between naming and category formation.

As a final observation, we remark that the sharpening of perceived between-category differences and attenuation of perceived within-category differences also known as categorical perception (CP) (37, 38), and observed in the CG dynamics, could be an innate property or an outcome of the process of language learning. In fact, there is a huge amount of literature in support of either of these conjectures. The former position can be historically connected to rationalism (39) and is often found either in an explicit or an implicit way in evolutionary psychology (40–42). Specific to colors, refs. 43 and 44 have tried to seek evidence toward a genetic coding of color categories by analyzing the color categorization behavior of newborn children. On the other hand, in support of the latter position, presence of learning has been demonstrated through color tests with prelanguage children (45–47) and by means of experiments where individuals from a particular culture were tasked to learn the color categories of another culture (18, 48). However, a majority of researchers agree that even learning-based induction of CP is "loosely constrained by the default neural organization," as has been suggested in ref. 18. The CG builds up on this last idea that the assumption of a minimal neural/physiological substrate (nonspecific to language) coupled with a complex cultural interaction process can actually cause the emergence of categorization patterns in a population of agents. It is important to note here that it is not only the neural substrate (i.e., JND) but also the complex dynamical process of learning of the agents that together lead to the observed hierarchy. In other words, the strong positive correlation between the JND and the hierarchical structure is not straightforward; in contrast, it is guided by a complex nonlinear chain of interactions.

Conclusion

In this paper, we have shown that a simple negotiation dynamics, driven by a weak nonlanguage specific bias, namely the frequency dependent resolution power of the human eye, is sufficient to guarantee the emergence of the hierarchy of color names getting so arranged by the times needed for their fixation in a population. The observed hierarchy features an excellent quantitative agreement with the empirical observations, confirming that the theoretical modeling in this area has now attained the required maturity to make significant contributions to the ongoing debates in cognitive science. Our approach suggests a possible route to the emergence of hierarchical color categories: The color spectrum clearly exists at a physical level of wavelengths, humans tend to react most saliently to certain parts of this spectrum often selecting exemplars for them, and finally comes the process of linguistic color naming, which adheres to universal patterns resulting in a neat hierarchy of the form obtained here. These intuitions are of course not a novelty (see for instance ref. 19); however, we provided a theoretical framework where the origin of the color hierarchy, as well as its quantitative structure, could be explained and reproduced through a purely cultural route driven, on its turn, by a nonlanguage-specific property of human beings.

It should be remarked that, despite the striking universal character of the color hierarchy, fluctuations exist across different languages as for the precise order in which color names got fixed in each language. In the framework of our model, this phenomenon is naturally explained as a consequence of the unavoidable stochasticity of the underlying cultural negotiation dynamics (15). The error bars in the fixation time of each specific color term in Fig. 4 specifically support this picture. Finally, it is important to mention that our results are paving the way for a detailed comparison with true historical data for each attested language, taking into account for instance phenomena like language contact and multilingualism as well as more language-specific cultural evolution processes.

Methods

The Category Game. The CG (15) constitutes of a set of N artificial agents in a simulated population with no words or categories at all in the beginning. As the game proceeds, the agents are repeatedly tasked with describing different perceptual stimuli received from their environment (e.g., colors) to one another. While doing so, a single stimulus (corresponding to a real value in a continuous perceptual space, e.g., the visible light spectrum) is chosen from a set of multiple such stimuli (named objects) present in the environment and is denoted as the topic to be described. Each game is played by a pair of agents where one of them acts as a speaker, trying to describe the topic by a name, while another acting as a hearer, trying to guess just by listening to the name which object the speaker is referring to. The individual agents independently invent words and categories and, based on the success or failure of their communications, adjust their own categories and vocabularies to increase the success in communication. A communication is deemed successful if the word the speaker used appeared in the hearer's vocabulary and allowed the hearer to identify the object the speaker meant. Further, the agents are endowed with a real property of human vision—i.e., the Just Noticeable Difference (Fig. 1)—by virtue of which they are not required to distinguish between those hues that a human eye cannot tell apart. In the following, we present a brief description of the important components of this model referring the reader to the Supporting Information for a more detailed description accompanied by a suitable illustration of the individual steps of the game (see the *SI Text*).

Basic dynamics. The population consists of N artificial agents each of them having a one-dimensional continuous perceptual space spanning, without any loss of generality, the $[0, 1)$ interval. Categorization simply corresponds to the partitioning of this space into discrete subintervals, which we shall call perceptual categories from now onward. Starting from a blank slate, each agent progressively develops a dynamical inventory of form-meaning associations linking categories (meanings) to words (forms). The emerging categories as well as the words associated to them coevolve over time through a series of simple communication interactions (or "games").

Choice of individuals for a game. In a game, two individuals are randomly selected from the set of N agents. One of them acts as a speaker and the other as a hearer. Both the speaker and the hearer are presented with a scene of $M \geq 2$* stimuli (objects), where each stimulus corresponds to a real number in the $[0, 1)$ interval. By definition, no two stimuli appearing in the same scene can be at a distance closer than $d_{\min}(x)$, where x can be either of the two. This function is the only parameter of the model encoding the finite resolution power of any perception or equivalently the human JND (Fig. 1).

Rules of negotiation. One of the objects is randomly denoted as the topic of the communication. This information is known only to the speaker. The task of the speaker is to communicate this information to the hearer using the following rule. The speaker always checks whether the perceptual category (i.e., the subinterval) in which the topic falls is unique for it. If the two stimuli fall in the same single perceptual category, then a new boundary is created in the perceptual space at a location corresponding to the middle of the segment connecting the two stimuli creating two smaller subintervals. A new name is invented for each of these two new perceptual categories. In addition, both of them inherit all the words corresponding to the old category. This process is termed as discrimination. Subsequently, the speaker utters the "most relevant" name for the category corresponding to the topic. The most relevant name is either the one used in a previous successful communication or the newly invented name in case the category has just been created due to a discrimination. For the hearer, there can be the following possibilities: (*i*) the hearer does not have any category associated with the name, in which case the game is a failure, or (*ii*) more than one categories are associated with this name in the hearer's inventory. In this case, the hearer randomly chooses one of them. If the hearer chooses the category linked to the topic, the game is a success, otherwise it is a failure.

Update of inventories. Depending on the outcome of the game, one or both the agents update their repertoires. In case of a failure, the hearer adds the word in her repertoire linked to the category corresponding to the topic. In case of a success, this word becomes the most relevant name for the category corresponding to the topic for both agents and they remove all the other competing words from their respective repertoires linked with this category.

Dynamical evolution. The dynamical evolution is initially driven by the pressure of discrimination, which makes the number of perceptual categories increase. At the same time, a synonymy emerges such that many different words are used by different agents for some similar categories. This kind of synonymy reaches a peak after which it starts to diminish as in the simple Naming Game (22). When on average only one word is recognized by the whole population for each perceptual category, a second phase of the evolution intervenes. During this phase, words expand their reference across adjacent perceptual categories, joining these categories to form the so-called linguistic categories. The coarsening of these categories features a dynamic arrest analogous to the physical process in which supercooled liquids approach the glass transition (26). On this long-lived state, the number of linguistic categories turns out to be finite and small (15).

Multilevel Emergence. Consider case *ii* discussed above. After the speaker transmits the name for the topic, the hearer finds more than one category associated with this name. If one of these categories is linked to the object and the hearer randomly chooses this one rather than the one linked to the topic, then we refer to this special case by failure with name. Note that this event is really not a "true" failure because, the hearer already knew the correct name for the topic and it can be associated to a confusion to differentiate between the topic and the object. We propose to overcome this situation by creating additional levels of linguistic categories. One can relate this scenario to the linguistic community of a set of specialized individuals, for instance, painters, for whom knowing the basic color terms are not enough to reach a reasonable communicative success. In contrast, knowledge of complex color terms are necessary to execute successful communication. Therefore, we include the possibility of creation of additional levels of linguistic categories in the CG model. We shall refer to the level corresponding to the basic CG as level 0 and the subsequent levels as level 1, level 2, and so on.

The precise prescription when a failure with name takes place is the following. A higher level is accessed if the entire range of perceptual categories between the one associated with the object and the one associated with the topic has the same name. Otherwise, the failure with name procedure does not apply. In the positive case, when an higher level is accessed, the procedure is as follows:

i. Activity of the hearer: In case a higher level does not exist yet, the hearer creates a new virtual level by filling the span of (*a*) all the perceptual categories (if any) that are adjoint to either the topic or the object and have the same name, (*b*) the perceptual categories corresponding to the topic and the object, and (*c*) all the intermediate perceptual categories between the topic and the object. This span is then divided into two parts with a boundary, borrowed from the $[0, 1)$ perceptual space, that corresponds (or is closest) to the midpoint of the two objects. The two parts of this virtual span are named by two brand new words.
ii. Activity of the speaker: Same as the activity of the hearer in *i*.
iii. Deletion of a span: If a higher level already exists for either of the agents, then they check whether the span filling (*a*), (*b*), and (*c*) (see step *i*) in the current level has an equal number of linguistic categories as the immediate lower level and, if so, this span is deleted.
iv. Game in the higher level: *If either* the conditions illustrated in steps *i* and *ii* are satisfied *or* the higher level for both the hearer and the speaker exists with at least a span filling the perceptual categories corresponding to the topic and the object as well as all the intermediate perceptual ca-

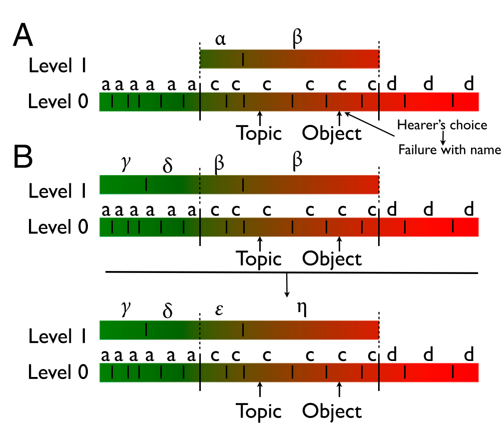

Fig. 5. Configuration of the different levels of a hypothetical agent. (*A*) The failure with name causes the creation of level 1 with two brand new words where the boundary is borrowed from level 0. (*B*) The number of linguistic categories in the span corresponding to the topic and the object is equal in level 0 and level 1 thereby causing a deletion of this span in level 1 followed by a recreation.

*Without any loss of generality in all our simulations, we shall use $M = 2$.

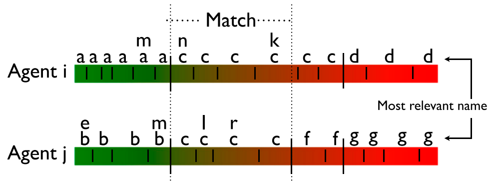

Fig. 6. Match between a pair of agents *i* and *j*. Note that for a match it suffices to have only the most relevant name similar in a particular region for an agent pair.

tegories between the topic and the object, *then* the speaker transmits the most relevant name corresponding to the topic, selecting this name now from the higher level inventory, and the game in this level continues following exactly the rules of the basic CG. At the end of the game, in case of a failure with name in this level, steps *i–iv* are repeated to create an even higher level.

In Fig. 5 *A* and *B*, we illustrate one representative example of the process of creation and deletion of spans in the higher level.

Match. A match region match(i, j) for a pair of agents *i* and *j* is the sum of the length of all the regions in their perceptual space where both of them have the same most relevant name. For instance, in Fig. 6, the match region corresponds to that length where both the agents have the same most relevant name "c." Note that this metric is a quantitative measure of the amount of agreement between the agent pair. The match of the whole population is simply

$$\frac{2\sum_{i=1}^{N}\sum_{j=i+1}^{N}\mathrm{match}(i,j)}{N(N-1)}. \quad [1]$$

ACKNOWLEDGMENTS. The authors are grateful to Andrea Baronchelli and Andrea Puglisi for very interesting and stimulating discussions. V.L. and F.T. acknowledge support from the European Specific Targeted Research Projects project EveryAware (Grant 265432).

1. Lakoff G (1987) *Women, Fire, and Dangerous Things: What Categories Reveal About the Mind* (Univ Chicago Press, Chicago).
2. Taylor J, Taylor J (2003) *Linguistic Categorization* (Oxford Univ Press, New York).
3. Murphy G (2004) *The Big Book of Concepts* (MIT Press, Cambridge, MA).
4. Whorf B (1956) *Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf*, ed JB Carroll (MIT Press, Cambridge, MA).
5. Chomsky N (1980) *Rules and Representations* (Columbia Univ Press, New York).
6. Berlin B, Kay P (1969) *Basic Color Terms* (Univ California Press, Berkeley, CA).
7. Ratliff F (1976) On the psychophysiological bases of universal color terms. *Proc Am Philos Soc* 120(5):311–330.
8. Kay P, Maffi L (1999) Color appearance and the emergence and evolution of basic color lexicons. *Am Anthropol* 101:743–760.
9. Kay P, Regier T (2003) Resolving the question of color naming universals. *Proc Natl Acad Sci USA* 100:9085–9089.
10. Regier T, Kay P, Khetarpal N (2007) Color naming reflects optimal partitions of color space. *Proc Natl Acad Sci USA* 104:1436–1441.
11. Cook R, Kay P, Regier T (2005) *The World Color Survey Database: History and Use*, eds H Cohen and C Lefebvre (Elsevier, Amsterdam), pp 224–242.
12. Steels L, Belpaeme T (2005) Coordinating perceptually grounded categories through language: A case study for colour. *Behav Brain Sci* 28(4):469–489.
13. Dowman M (2007) Explaining color term typology with an evolutionary model. *Cogn Sci* 31:99–132.
14. Komarova NL, Jameson KA, Narens L (2007) Evolutionary models of color categorization based on discrimination. *J Math Psychol* 51:359–382.
15. Puglisi A, Baronchelli A, Loreto V (2008) Cultural route to the emergence of linguistic categories. *Proc Natl Acad Sci USA* 105:7936–7940.
16. Baronchelli A, Gong T, Puglisi A, Loreto V (2010) Modelling the emergence of universality in color naming patterns. *Proc Natl Acad Sci USA* 107:2403–2407.
17. Saunders B, Van Brakel J (1997) Are there nontrivial constraints on colour categorization? *Behav Brain Sci* 20:167–179.
18. Roberson D, Davies I, Davidoff J (2000) Color categories are not universal: Replications and new evidence from a stone-age culture.. *J Exp Psychol Gen* 129:369–398.
19. Evans N, Levinson SC (2009) The myth of language universals: Language diversity and its importance for cognitive science. *Behav Brain Sci* 32:429–448.
20. Wittgenstein L (1953) *Philosophical Investigationstrans Anscombe GEM* (Basil Blackwell, Oxford).
21. Steels L (1995) A self-organizing spatial vocabulary. *Artif Life* 2:319–332.
22. Baronchelli A, Felici M, Caglioti E, Loreto V, Steels L (2006) Sharp transition towards shared vocabularies in multi-agent systems. *J Stat Mech* P06014.
23. Komarova NL, Jameson KA (2008) Population heterogeneity and color stimulus heterogeneity in agent based color categorization. *J Theor Biol* 253:680–700.
24. Komarova NL, Jameson KA (2009) Evolutionary models of color categorization. I. Population categorization systems based on normal and dichromat observers. *J Opt Soc Am A* 26:1414–1423.
25. Komarova NL, Jameson KA (2009) Evolutionary models of color categorization. II. Realistic observer models and population heterogeneity. *J Opt Soc Am A* 26:1424–1436.
26. Mukherjee A, Tria F, Baronchelli A, Puglisi A, Loreto V (2011) Aging in language dynamics. *PLoS One* 6:e16677.
27. Bedford RE, Wyszecki G (1958) Wavelength discrimination for point sources. *J Opt Soc Am* 48:129–135.
28. Long F, Yang Z, Purves D (2006) Spectral statistics in natural scenes predict hue, saturation, and brightness. *Proc Natl Acad Sci USA* 103:6013–6018.
29. Plümacher M, Holz P, eds. (2007) *Speaking of Colors and Odors* (John Benjamins Publishing, Amsterdam).
30. Kay P, McDaniel C (1978) The linguistic significance of the meanings of basic color terms. *Language* 54:610–646.
31. Abramov I (1997) *Physiological Mechanisms of Color Vision*, eds CL Hardin and L Maffi (Cambridge Univ Press, Cambridge, UK), pp 89–117.
32. Berk T, Kaufman A, Brownston L (1982) A human factors study of color notation systems for computer graphics. *Commun ACM* 25:547–550.
33. Lucy J (1992) *Grammatical Categories and Cognition: A Case Study of the Linguistic Relativity Hypothesis* (Cambridge Univ Press, Cambridge, UK).
34. Lucy J (1992) *Language Diversity and Thought. A Reformulation of the Linguistic Relativity Hypothesis* (Cambridge Univ Press, Cambridge, UK).
35. Imai M, Gentner D (1997) A cross-linguistic study of early word meaning: Universal ontology and linguistic influence. *Cognition* 62:169–200.
36. Lucy J, Gaskins S (2001) *Language Acquisition and Conceptual Development*, eds M Bowerman and S Levinson (Cambridge Univ Press, Cambridge, UK), pp 257–283.
37. Harnad S, ed. (1990) *Categorical Perception: The Groundwork of Cognition* (Cambridge Univ Press, Cambridge, UK).
38. Harnad S (2006) Categorical Perception. *Encyclopedia of Cognitive Science* (Nature Publ Group: Macmillan, New York).
39. Fodor JA (1983) *The Modularity of Mind* (MIT Press, Cambridge, MA).
40. Pinker S, Bloom P (1990) Natural languages and natural selection. *Behav Brain Sci* 13:707–784.
41. Durham WH (1991) *Coevolution: Genes, Culture and Human Diversity* (Stanford Univ Press, Palo Alto, CA).
42. Shepard RN (1994) Perceptual-cognitive universals as reflections of the world. *Psychon Bull Rev* 1:2–28.
43. Bornstein MH, Kessen W, Weiskopf S (1976) Color vision and hue categorization in young human infants. *J Exp Psychol Hum Percept Perform* 2:115–129.
44. Gerhardstein P, Renner P, Rovee-Collier C (1999) The roles of perceptual and categorical similarity in colour pop-out in infants. *Br J Dev Psychol* 17:403–420.
45. Bornstein MH (1975) The influence of visual perception on culture. *Am Anthropol* 77:774–798.
46. Davies I, Franklin A (2002) Categorical perception may affect colour pop-out in infants after all. *Br J Dev Psychol* 20:185–203.
47. Franklin A, et al. (2008) Categorical perception of color is lateralized to the right hemisphere in infants, but to the left hemisphere in adults. *Proc Natl Acad Sci USA* 105:3221–3225.
48. Rosch-Heider E (1972) Universals in color naming and memory. *J Exp Psychol* 93:10–20.