PLOS | ONE

RESEARCH ARTICLE

# Individual Biases, Cultural Evolution, and the Statistical Nature of Language Universals: The Case of Colour Naming Systems

**Andrea Baronchelli**[1]*, **Vittorio Loreto**[2,3,4], **Andrea Puglisi**[5,2]

**1** Department of Mathematics, City University London, London EC1V 0HB, UK, **2** Dipartimento di Fisica, Sapienza Università di Roma, P.le A. Moro 5, 00185 Roma, Italy, **3** Institute for Scientific Interchange (ISI), Via Alassio 11/c, 10126, Turin, Italy, **4** SONY Computer Science Lab (SONY-CSL), 5, Rue Amyot, 75005, Paris, France, **5** CNR-ISC, Piazzale Aldo Moro 5, 00185 Roma, Italy

* a.baronchelli.work@gmail.com





**Citation:** Baronchelli A, Loreto V, Puglisi A (2015) Individual Biases, Cultural Evolution, and the Statistical Nature of Language Universals: The Case of Colour Naming Systems. PLoS ONE 10(5): e0125019. doi:10.1371/journal.pone.0125019

**Academic Editor:** Kenny Smith, University of Edinburgh, UNITED KINGDOM

**Received:** October 13, 2014

**Accepted:** March 8, 2015

**Published:** May 27, 2015



**Data Availability Statement:** All relevant data are within the paper.

**Funding:** The authors are grateful to Animesh Mukherjee and Francesca Tria for helpful discussions. V.L. acknowledges support from the EveryAware European project nr. 265432 under FP7-ICT-2009-C and DRUST project funded by the European Science Foundation under EuroUnderstanding Collaborative Research Projects and the KREYON project funded by the Templeton Foundation under contract N. 51663. SONY-CSL provided support in the form of salaries for author VL, but did not have any additional role in the study

## Abstract

Language universals have long been attributed to an innate Universal Grammar. An alternative explanation states that linguistic universals emerged independently in every language in response to shared cognitive or perceptual biases. A computational model has recently shown how this could be the case, focusing on the paradigmatic example of the universal properties of colour naming patterns, and producing results in quantitative agreement with the experimental data. Here we investigate the role of an individual perceptual bias in the framework of the model. We study how, and to what extent, the structure of the bias influences the corresponding linguistic universal patterns. We show that the cultural history of a group of speakers introduces population-specific constraints that act against the pressure for uniformity arising from the individual bias, and we clarify the interplay between these two forces.

## Introduction

### Language universals and colour naming

Different languages share a collection of structural properties, which are accordingly said to be *universal*[1]. The origin and nature of this universality have been debated for decades in a dispute that is still far from being settled [2, 3]. The classical view assumes that language learning requires innate language-specific knowledge [4, 5]. Every individual is endowed with a set of grammatical principles, defining a Universal Grammar, that are consequently found to be shared across all human languages. This view has however been questioned in various ways (see, for example, [6]), and different computational approaches have shown that the hypothesis of a language-specific genetic endowment would result in a series of paradoxes [7], while cultural transmission introduces informational bottlenecks that favour the emergence of regularities [8, 9]. A systematic analysis has also shed new light on the very concept of universality

design, data collection and analysis, decision to publish, or preparation of the manuscript. The specific roles of the authors are articulated in the "author contributions" section. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

[10]. Language universals must be intended in a statistical sense, rather than as necessary features of all human languages [6, 10].

All of these more recent accounts, although heterogenous in other respects, agree that the observed cross-linguistic regularities emerged independently in every language in response to universal communicative, cognitive, or perceptual biases that are not specific to language [11]. However, several issues remain open. Evans and Levinson [10] have recently listed some of the more urgent questions, among which are: What are the biases responsible for a *given* observed regularity? How do they generate a structure? What is the relation between the strength of a bias and the amount of cross-linguistic variability or universality? To this list we would add a question which is to us as urgent as the others: what is the origin of cross-linguistic variability [12]? In other words, how effective is the specific history of a language in shaping specific, non universal properties?

Answering these questions is hard, but focusing on simple aspects of language has provided important insights. This is why colour categorisation has captured a great deal of attention over the last few decades. As observed in the World Color Survey (WCS) [13, 14], different groups of individuals develop different colour naming patterns, but some universal properties can be identified by a statistical analysis over a large number of populations [15]. Furthermore, it appears that basic colour names enter a language in a relatively fixed order [13]. In general, the universality of colour naming is fully recognised as a genuine linguistic universal [11, 16–21], and has been the focus of many research efforts. In particular, several computational models have contributed to shedding light on the different theoretical hypotheses put forward in this context [22–33]. Even though most of these efforts have been devoted to modelling the emergence of shared colour categorisation patterns in a single population of individuals, more recent contributions have also addressed the issue of universality [24, 30, 31, 34].

## The Category Game model

Here, we focus on the Category Game model [27], describing how a group of agents manages to establish a shared set of linguistic categories when faced with a continuous aspect of the environment, such as the hue channel of the colour spectrum. According to the standard approach of language games [35, 36], a population of agents can establish a shared category system through pairwise linguistic interactions alone. In every conversation one of the two agents tries to direct the attention of the other towards a specific object out of those that constitute the scene they are both facing. Depending on the success (or failure) of this operation both agents modify their internal state eventually leading to a global consensus. The model can be informed with a human psychophysiological parameter, namely the Just Noticeable Difference (JND), defining the minimum distance at which two stimuli from the same scene can be discriminated as a function of their wavelength [37, 38]. If the individuals are endowed with the human JND function, the statistical properties of the emerging categorisation pattern turn out to quantitatively match those observed in the World Color Survey data [30], and the empirically observed colour hierarchy is reproduced [31].

The agreement between the outcome of the model and experimental data corroborates the hypothesis according to which weak perceptual or cognitive biases can be responsible for the emergence of linguistic properties shared across independent languages [6, 10, 11]. However, in the model the bias does not constrain the structure of the shared categorisation pattern in a deterministic way. In fact, two forces shape the naming structure of a single population. Cultural history defines arbitrary consensus patterns that, once they have emerged, remain as frozen accidents and affect the further evolution of the language [39, 40], while the perceptual bias tends to prefer some patterns over others. At the level of a single population it is not possible to

predict which force will prevail and to what extent. Statistically, however, it is possible to recover the signature of the bias.

Here, we investigate the role of the individual bias. We clarify, in the framework of the model, what a "weak" bias is, how it works, and how it affects the cultural history of a linguistic pattern. To do so we exploit the flexibility offered by numerical simulations in two ways. First, we manipulate the nature of the bias, considering artificial JND functions, and comparing the impact they have on the emerging categorisation patters. We find that the human JND belongs to a class of special functions that allow for a good agreement with the experimental data, and that this agreement is by no means assured by any arbitrary JND. Then, we analyse in detail what "weak" bias means by looking at the variability of the observed patterns. In particular, we study the cross-linguistic fluctuations of the categories which allow us to estimate the relative strength of cultural and cognitive or perceptual pressures in the framework of the model. We show that single languages can be driven far from universality by their own historical evolution, even though the vast majority of languages share, to some extent, universal properties. The peculiar nature of such fluctuations (characterised by *fat* tails, i.e., over-populated extreme events) suggests a relevant role for history-dependent correlations [40], which—in a sense—should be considered as a kind of *cultural pressure*.

## Related work

The role of weak biases has been investigated also in relation to their impact on cultural transmission [41]. In particular, Kirby et al. [8] have investigated how a biological bias can translate into universal linguistic properties in the context of the Iterated Learning framework [42], assuming that learners apply the principles of Bayesian inference [43]. The central result is that cultural transmission can magnify weak biases into strong universals, while strong biases could be shielded by transmission bottleneck, i.e., the limited amount of linguistic examples from which each speaker must learn the language. In the same framework, [9] showed that linguistic structure can emerge when language learners use learning algorithms only slightly biased towards structured languages. As we mentioned above, and will see in the remainder of the paper, our work nicely complements these results by showing that the history of each language can introduce non-biological constraints as further biases shaping the language itself. While the Iterated Learning approach sheds light on the interplay between biology and cultural transmission in a formal way, we tackle, from a different perspective, the relationship between biology and cultural evolution.

Another interesting result concerns a numerical experiment by Regier et al [44]. Here the hypothesis that colour naming represents optimal partitions of the colour space [45] is tested through the definition of a measure of "well-formedness" of the linguistic categories for colour. Optimal partitions of colour space are those that maximise this well-formedness, and the authors find that artificially generated colour-naming schemes that maximise this quantity in the full CIEL$^*$a$^*$b space resemble the colour naming scheme of some of the world's languages, even though "many languages" are not "very similar" to the hypothetically optimal model configurations. A second result is that given the colour naming scheme of any language, distortions away from that pattern often result in lower well-formedness. While this approach does not explain the observed number of categories, which is a parameter to be fixed before performing the analysis, it provides an interesting perspective on colour categorisation patterns. In particular, as the authors suggest, the fact that optimality does not account for all of the observed languages could be the result of cultural forces pushing languages far from the optimal partition. From this perspective, the Category Game helps to substantiate this argument by showing how culture and a universally shared parameter (here the candidate would be optimality) can interact.

More recently, Xu et al. have shown that cultural transmission produces schemes of colour categorisation similar to those observed in the WCS in a laboratory experiment in which chain of individuals simulated the Iterated Learning scheme with a pre-determined number of terms [34]. In particular, the authors borrow tools from information theory to show a significant match between their experiments and the WCS data. Also in this case, we believe that the Category Game represents a valuable complementary approach (i) offering a possible mechanism determining the cultural evolution of the small number of colour names observed in all the languages of the WCS, and (ii) suggesting with quantitative evidence that the underlying perceptual bias, magnified by cultural transmission in the Xu et al. experiment [34], could in fact be the human JND function.

Finally, among the different computational models set forth to study colour categorisation [22–33], it is important to mention the one introduced in [24], which replicates the frequency of colour terms observed in the WCS by adopting an evolutionary perspective, along with a Bayesian scheme of term acquisition, and dealing with a one-dimensional colour space relative to hue dimension (as the Category Game does). In [24] the hue space is discrete, featuring 40 possible hue values, and the model assumes (i) that the universal foci are predefined and unevenly spaced in the colour space and (ii) that colour terms denote a contiguous range of colours. In the Category Game, on the contrary, the hue space is continuous and the two hypotheses above are not necessary. In particular, the existence of universal foci is a byproduct of the fact that the individuals are endowed with the human JND function, while the fact that colour names refer to contiguous frequencies is an emerging property of the dynamics [27].

Considered together, the two models of [24] and [27] indicate that at least some properties of the WCS data can be accounted for by considering the hue channel alone. Specifically, this is true for basic colour terms, while it might not be the case for the naming of composite or derived colours [46]. Interestingly, the crucial role played by hue has recently found a biological rationalisation in the neurological study of primate vision. According to Xiao et al., the fact that the non-random properties of colour-naming patterns can be accounted for by the wavelength JND, as shown by [30], implies a link between the universal constraints and the functional characteristics of the retinogeniculate pathway [47]. And in fact, the study carried on by the same authors establishes a direct link between a universal constraint on colour naming, namely the fault line close to the "warm"-"cold" colour distinction, and the cone-specific information that is represented in the primate early visual system, specifically the L- versus M-cone contrast [47].

The Category Game has in fact been extended to more dimensions in order to describe the evolution from a mostly brightness-based to a mostly hue-based colour term system comparable to what happened during the Middle English period in response to the rise of dyeing and textile manufacturing [48]. The 2D model is obviously more complex and requires further specifications of the rules determining the individual categorisation process. The outcome is richer than its one-dimensional counterpart, but the unavoidable consequence is that it is less transparent to interpretation. Since our aim here is to clarify the issue of the interplay between a cognitive or perceptual bias and the universality properties of the generated languages, and that the latter happens to appear already in the 1D model, we stick to the simplest case (see [49] for a deeper discussion on why choosing simplicity when confronted with such a choice).

## Materials and Methods

The essential features of the Category Game model are:

1. It considers the categorisation of a continuum perceptual channel [27];

2. It generates categorisation patterns whose statistical properties are in quantitative agreement with the ones observed in the WCS data [30];

3. It reproduces the hierarchy of colour terms [31].

In this section we briefly recall the model definition and some previous results.

## The model

The Category Game involves a population of $N$ artificial agents. Starting from scratch and without any pre-defined colour categories, the model dynamically leads, through a sequence of pair-wise interactions (games), to the emergence of a highly shared set of linguistic categories of the visible light spectrum. The parameters of the model are the population size $N$ and the JND function.

For the sake of simplicity and without loss of generality (see also [48]), colour perception is reduced to a single analogical continuous perceptual channel, each light stimulus being a real number in the interval [0, 1), which represents a rescaled wavelength. A categorisation pattern is identified with a partition of the interval [0, 1) in sub-intervals, or perceptual categories. Individuals have dynamical inventories of form-meaning associations linking perceptual categories with their linguistic counterparts, basic colour terms, and these inventories evolve through elementary language games. At each time step, two players (a speaker and a hearer) are randomly selected from the population and a scene of $M > 1$ stimuli is presented. Two stimuli cannot appear at a distance smaller than JND($x$) where $x$ is the value of one of the two. This is the way in which the JND is implemented in the model. On the basis of the presented stimuli, the speaker discriminates the scene, refining if necessary its perceptual categorisation, and utters the colour term associated with one of the stimuli. The hearer tries to guess the named stimulus, and based on their success or failure, both individuals rearrange their form-meaning inventories. New colour terms are invented every time a new category is created for the purpose of discrimination, and are spread through the population in successive games. A detailed description of the model is presented in the Appendix.

The dynamics proceeds as follows [27]. At the beginning all individuals have only the perceptual category [0, 1) with no associated name. During the first phase of the evolution, the pressure for discrimination makes the number of perceptual categories increase, resulting in widespread synonymy due to the many different words used by different agents for similar categories. This kind of synonymy reaches a peak and then dies out, in a similar way as in the Naming Game [36, 50]. When on average only one word is shared by the whole population for each perceptual category, a second phase of the evolution starts. During this phase, words expand their dominion across adjacent perceptual categories, merging several perceptual categories giving rise to a new type of categories, namely the "linguistic categories". The structure of the linguistic categories evolves through a domain growth process, known in statistical physics as coarsening [51]. Their number is progressively reduced till the system undergoes a dynamical arrest, featuring a slowing down of the domain growth, along much the same lines as the physical processes by which supercooled liquids approach the glass transition [52].

In this long-living, almost stable, phase, usually after $10^4$ games per player, the linguistic categorisation patterns of the individuals overlap with each other to an extent ranging between 90% and 100%. The success rate and the similarity of category patterns across different agents remain stable for a time of $\sim 10^6$ games per player, and this pattern is considered as the final categorisation pattern generated by the model, to be compared with human colour categories (see below). Note that, at the level of the Category Game, categories can be equivalently described in terms of boundaries or prototypes. The shared pattern in the long stable phase

between $10^4$ and $10^6$ games per player is the main subject of the experiment described in the following section. As already observed in [27] the number of linguistic colour categories observed in this phase is of the order of 15±10, as in natural languages, even though discrimination allows for the existence of hundreds of categories.

## Comparison with real-world data

A large amount of data on colour categorisation was gathered in the World Color Survey [13, 15], in which individuals belonging to different cultures had to name a set of colours. In this section, we review how the Category Game model has been used to reproduce some of the most important features of the empirical data.

The fundamental source of experimental data on colour naming systems comes from the World Colour Survey. P. Kay and B. Berlin [13] ran a first survey on 20 languages in 1969. From 1976 to 1980, the enlarged World Color Survey was conducted by the same researchers along with W. Merrifield [53] and the data have been publicly available since 2003 on the website http://www.icsi.berkeley.edu/wcs. These data concern the basic colour categories in 110 languages without written forms and spoken in small-scale, non-industrialised societies. On average, 24 native speakers of each language were interviewed. Each informant had to name each of 330 colour chips produced by the Munsell Color Company that represent 40 gradations of hue and maximal saturation, plus 10 neutral colour chips (black-gray-white) at 10 levels of value. The chips were presented in a predefined, fixed random order, to the informant who had to tag each of them with a "basic colour term" in her language (for more details see [13]).

After two decades of intense debate [19], Kay and Regier [15] performed a quantitative analysis proving that the colour naming systems obtained in different cultures and language are in fact not random. Through a suitable transformation, they identified the most representative chip for each colour name in each language and projected it into a suitable metric colour space (namely, the CIEL*a*b* colour space). To investigate whether these points are more clustered across languages than would be expected by chance, they defined a dispersion measure on this set of languages $S_0$

$$D_{S_0} = \sum_{l,l^* \in S_0} \sum_{c \in l} \min_{c^* \in l^*} \text{distance}(c, c^*), \tag{1}$$

where $l$ and $l^*$ are two different languages, $c$ and $c^*$ are two basic colour terms respectively from these two languages, and distance($c$, $c^*$) is the distance between the points in colour space in which the colours are represented. To give a meaning to the measured dispersion $D_{S_0}$, Kay and Regier created "new" datasets $S_i$ ($i$ = 1, 2, .., 1000) obtained through random rotations of the original set $S_0$, and measured the dispersion of each new set $D_{S_i}$. The human dispersion appears to be distinct from the histogram of the "random" dispersions with a probability larger than 99.9%. As shown in Fig. 3A of [15], the average dispersion of the random datasets, $D_{neutral}$, is 1.14 times larger than the dispersion of human languages. Thus, human languages are more clustered, i.e., less dispersed, than their random counterparts, confirming the existence of some kind of universality.

Another important finding concerning the World Colour Survey is the existence of a hierarchy of basic colour names that were adopted by individual cultures in a relatively fixed order [13]. Thus, basic colour terms can be organised into a hierarchy around the focal colours black, white (where black and white usually map 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), red, green, yellow, and blue always appearing in this order across cultures in such a way that if a culture has, for example a word for red, it also has a name for black

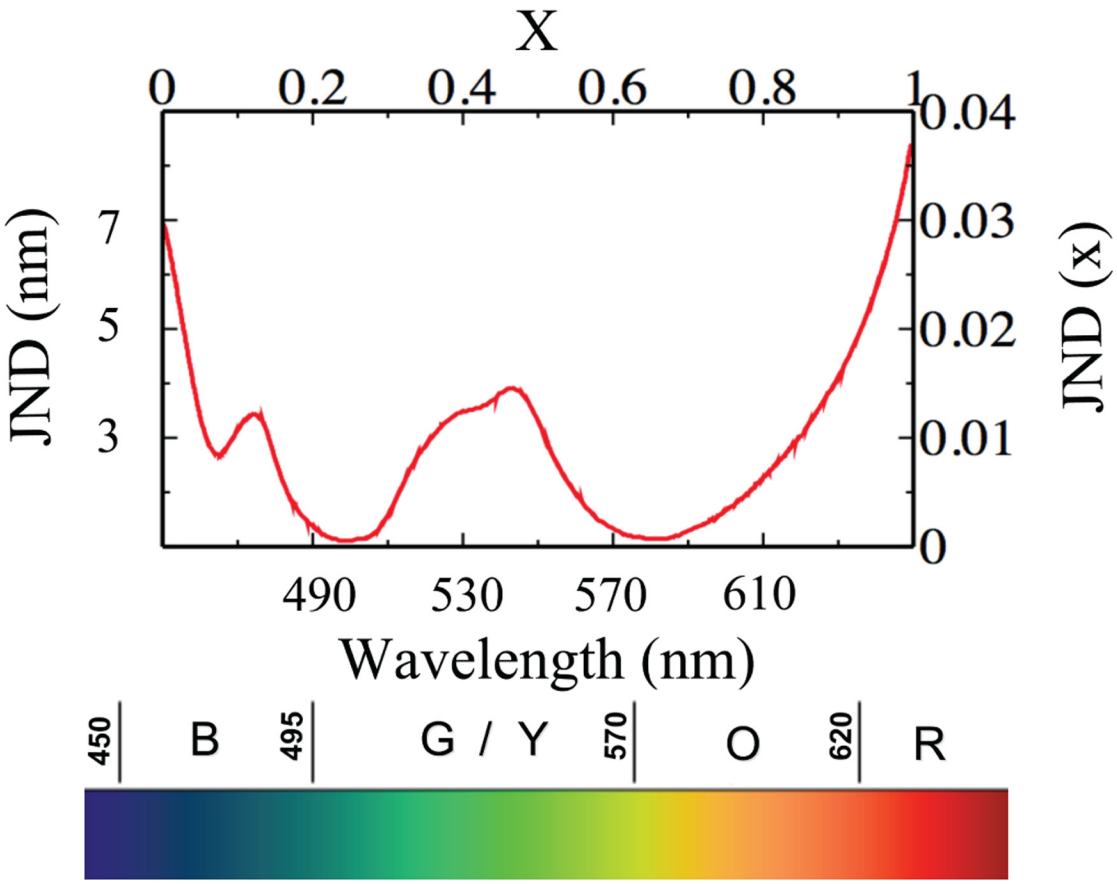


**Fig 1. The human Just Noticeable Difference (JND) function describes the wavelength change in a monochromatic stimulus needed to elicit a particular JND in the hue space.** Both as a function of the wavelength of the incident light (measured in nanometers) and on the rescaled interval [0, 1). For convenience we display also the spectrum of the visible light. For the purpose of the Category Game we rescale the monochromatic stimulus in visible spectrum (measured in nanometers) in the range [0, 1) for the Topic *x*. In the same way the JND function (in nanometers in the left y-axis) is rescaled into a *JND*(*x*) function (right y-axis).

doi:10.1371/journal.pone.0125019.g001

and white (but not vice-versa), if it has a name for green it also has a name for red (but not vice-versa) etc.

**The Numerical World Color Survey.** The core of the analysis described above is the comparison of the clustering properties of a set of *true* human languages against the ones exhibited by a certain number of randomised sets. In replicating the experiment it is therefore necessary to obtain two sets of synthetic data, one of which must have some human ingredient in its generation. The idea put forth in [30] is to act on the JND function. Human beings are endowed with a JND for the hue that is a function of the wavelength of the incident light (see Fig 1). This is the only parameter of the model encoding the finite resolution power of perception, or equivalently the human Just Noticeable Difference.

Starting from the human JND, different artificial sets can be created:

- *"Human"* categorisation patterns are obtained from populations whose individuals are endowed with the rescaled human JND;

- *Neutral* categorisation patterns are obtained from populations in which the individuals have constant JND (JND = 0.0143), which is the average value of the human JND (as it is projected on the [0, 1) interval).

In analogy to the WCS experiment, the randomness hypothesis in the NWCS for the neutral test-cases is supported by symmetry arguments: in neutral simulations there is no breakdown of translational symmetry in the colour space, which is the main bias in the"human" simulations.

Thus, the difference between "human" and neutral data originates from the perceptual architecture of the individuals of the corresponding populations. A collection of "human" individuals form a "human" population, and will produce a corresponding "human" categorisation pattern. In a hierarchical fashion, finally, a collection of populations is called a *world*, which in [30] is formed either by all "human" or by all non-"human" populations. For each world, the value of the dispersion $D$, defined in Eq (1), is calculated, in order to quantify the amount of dispersion of the languages (or categorisation patterns) belonging to it. In the actual WCS there is of course only one human World (i.e., the collection of 110 experimental languages), while in [30] several (i.e., 1500) worlds were generated to gather statistics both for the "human" and non-"human" cases.

The main results of the NWCS is that the Category Game, informed with the human *JDN* ($x$) curve, produces a class of "worlds" featuring a dispersion lower than and well distinct from that of the class of "worlds" endowed with a non-human, i.e., uniform, $JND(x)$. Moreover, the ratio observed in the NWCS between the average dispersion of the "neutral worlds" and the average dispersion of the "human worlds" is $D_{neutral}/D_{human} \sim 1.14$, very similar to the one observed between the randomised datasets and the original experimental dataset in the WCS. Crucially, these findings are robust against changes in such parameters as the population size $N$, the distribution of the stimuli, the number of objects in a scene $M$, and the time of measurement (as long as measures are taken in the temporal region in which a categorisation pattern exists) [30].

It is worth noting that often the outcome of a multi-agent model is employed to test the viability of theoretical hypotheses at a more abstract level [49]. In this case, the results of the NWCS compare with WCS data not only from a qualitative point of view, but also quantitatively. Moreover, the very design of the model suggests a possible mechanisms lying at the roots of the observed universality, previously formulated on the basis of theoretical analysis (see, for instance [6, 11]). Human beings share certain perceptual biases that, even though not so strong as to deterministically affect the outcome of the categorisation process, are capable of influencing category patterns. However, this influence can be made evident only through a statistical analysis performed over a large number of languages, since any two (or few) languages could provide a misleading signal, either in the direction of strong universality or on the contrary of a total lack of shared features.

**Reproducing the hierarchy of colour names.** To address the problem of the origin of a hierarchy of colours the Category Game has been generalised to allow for the emergence of a series of shared linguistic layers each of which could guarantee communicative success in progressively more complex tasks [31]. In the generalised version a higher and more refined linguistic layer is accessed by the agents only if in a game both the topic and the object have the same name and thus, there is a 'failure' to differentiate between the two, possibly resulting in a failure in communication, referred as "failure with name". While the first layer of linguistic categories (Level 0, the one present also in the original Category Game) can be likened to the emergence of primary colour names, the successive layers (Levels 1, 2, . . .) might be linked to the emergence of complex colour names when the knowledge of the primary colour names is

not enough to achieve a reasonable communicative success (one can think of a linguistic community comprising specialised individuals, for instance painters, textile and cosmetic manufacturers [54]). Even though a detailed discussion of this generalisation of the Category Game is out of the scope of this paper (we refer for this to [31]), it is interesting to summarise the main results.

A first observation concerns the frequency of access to higher levels of linguistic categorisation as a function of the local value of the JND. It turns out that the agents need to access the higher level early in regions (in the hue space) corresponding to high values of the JND, while they access it quite late in regions corresponding to low values of the JND. This indicates that an agreement at Level 0 is reached faster in regions with high values of the JND resulting in more cases of "failure with name" in these regions, thereby, forcing the agents to access Level 1. Starting from this observation it is interesting to compute the time needed to get consensus in different regions of the hue space corresponding to different values of the JND. It turns out that the emergence of consensus occurs first in regions corresponding to high values of the JND while it occurs last in regions corresponding to very low JND. Strikingly, if the regions are arranged according to the time to reach a desired level of consensus, then they get organised into a hierarchy with [red, (magenta)-red], [violet], [green/yellow], [blue], [orange] and [cyan] (or [cyan] and [orange] as is usually observed for secondary basic colour names) appearing in this order. The names on which faster agreement is reached turn out to be the basic colour names that first emerge in a population, in excellent quantitative agreement with the empirical observations of the World Colour Survey. While we will not use the generalised model in this paper, this finding confirms the validity of the Category Game framework and and its ability to compare to and reproduce the empirical data of the WCS.

## Results

In this section we investigate the role of the perceptual bias. First, we check to what extent the very good agreement between the outcome of the model and the WCS data can be related to the specific structure of the human JND. We show that different instantiations of the JND bias significantly alter the agreement between simulations and true data. This result rules out the possible criticism that any kind of bias would lead to the same kind of universality in the model.

Then, we compare the categorisations of different simulated populations, and we measure the dispersion of their colour naming patterns in order to study the interplay between the bias and the stochasticity introduced by the underlying cultural process. We show that the stochastic cultural process can generate shared category patterns that affect the ensuing evolution, effectively competing with the cross-cultural ordering tendency of the bias. This explain why different populations can present significantly different categorisation patterns even when all individuals share the same JND function. We argue that our results allow us to better understand what a "weak" bias is, and to what extent it may affect the emergence of the universal properties of language.

### Artificial JNDs

Only two JND functions have been tested in previous works, namely the human and the flat uniform JNDs, in order to contrast the naturalistic case with an unbiased neutral one. Here, we exploit the opportunity offered by numerical simulations to test different, artificial, JNDs. The aim of this new experiment is to understand whether the agreement with data is related to the specific shape of the human JND or rather if it can be achieved, in the framework of the model, with any JND. Of course, any bijettive function defined in [0, 1) could play the role of a JND,

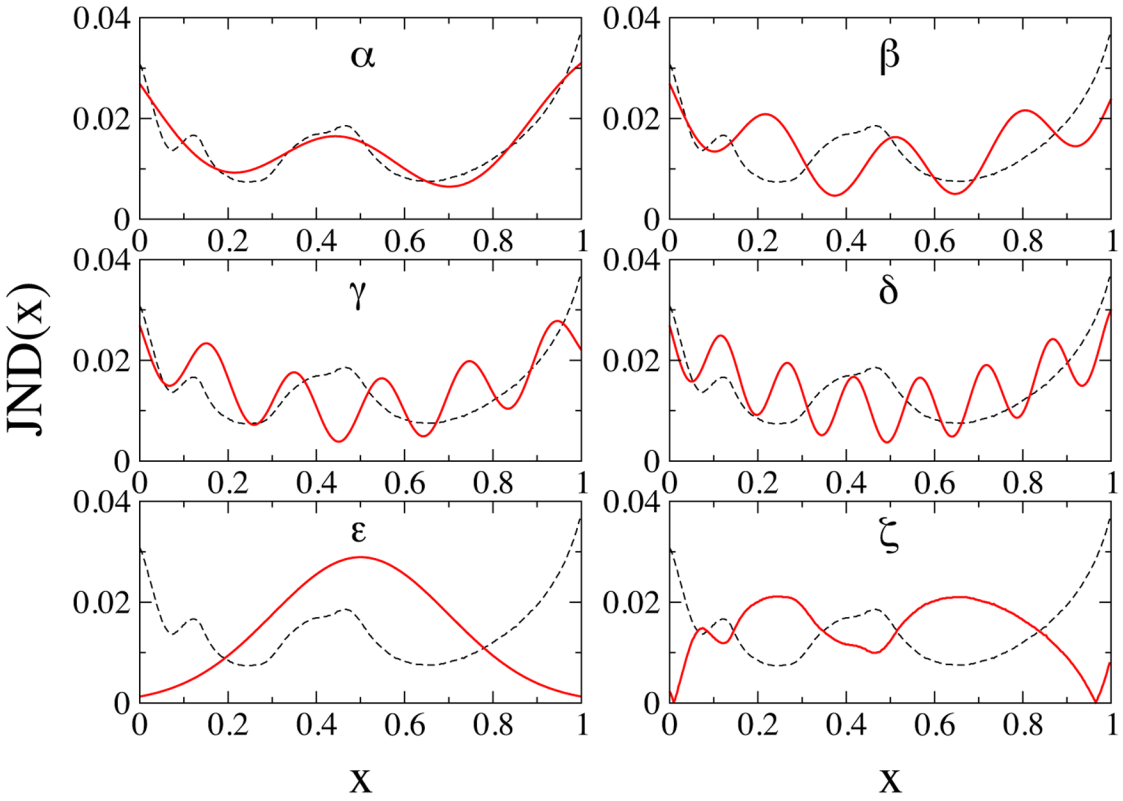


**Fig 2. Artificial JNDs.** Each panel depicts one of the artificial JNDs used in the experiments (continuous lines). The empirical fit of the human JND (dashed line, all panels) is obtained according to the expression $JND(x) = c_1 + c_2\cos(ax + c_3) + c_4(x - 0.5)^2$, where $c_1$, $c_2$, $c_3$, $c_4$ and $a$ are fitting constants (case $\alpha$). By constant increases of the $a$ parameter one finds increasingly irregular artificial JNDs ($\beta$, $\gamma$ and $\delta$). The last two panels present a Gauss-like JND ($\epsilon$), and a reflection of the JND around the average value (with a further prescription avoiding negative values of the JND) ($\zeta$).

doi:10.1371/journal.pone.0125019.g002

and a systematic exploration is therefore unfeasible, nor would it be useful. Rather, we start from the human JND, fit it empirically, and then modify the fitted function so as to progressively diverge from the original one. In particular, we insert local extreme points (minima and maxima) to increase the roughness of the JND. Furthermore, we consider also a Gauss-shaped JND as well as an inverted human JND. Fig 2 sketches the adopted functions. The key quantity studied here is, again, the dispersion $D$ defined in Eq (1) among languages, which is the natural measurement (complementary) of the degree of similarity, as already discussed in the literature [15].

Fig 3 shows the outcome of the numerical experiments run with several artificial JNDs. It is clear that the human JND, along with slight variations of it, performs well, while more irregular functions (obtained according to the expression $JND(x) = c_1 + c_2\cos(ax + c_3) + c_4(x - 0.5)^2$, where $c_1$, $c_2$, $c_3$, $c_4$ and $a$ are fitting constants) weaken the agreement with the empirical data (Cases $\beta$, $\gamma$ and $\delta$ are obtained by constant increases of the $a$ parameter). The Gaussian-like function (case $\epsilon$) and the inverse JND (case $\zeta$), perform worse than the human JND even thought not dramatically badly. It is worth stressing that the relative error between the results obtained with the simulated JND and the experimental one of [15] ($e = |sim - exp|/exp$) is plotted in logarithmic scale (Fig 3, bottom) and that an error below $e = 10\%$ can very well be considered as noticeably small given the simplicity of the Category Game model.

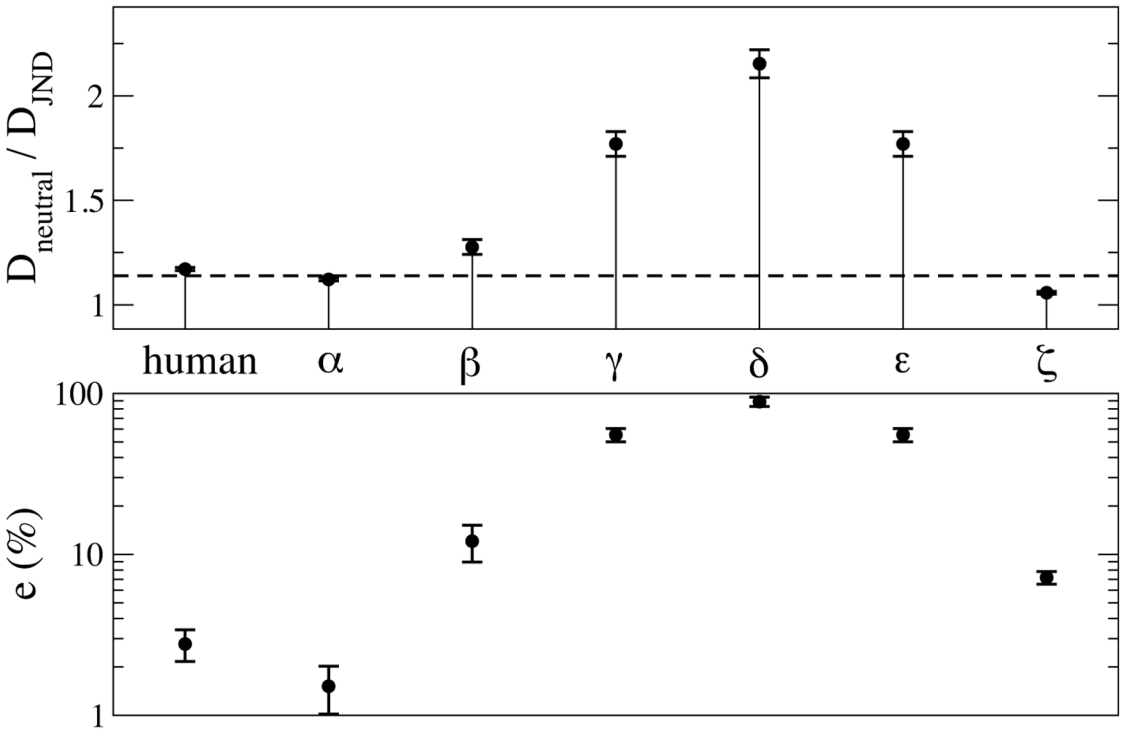


**Fig 3. Numerical World Color Survey with artificial JNDs.** Top: the dispersion of the JND case, as defined in Eq 1, normalised by the neutral one obtained with a flat JND is plotted for populations of size $N = 100$. The horizontal line indicates the experimental value 1.14 obtained by the analysis of WCS data in [15]. Bottom: the relative error $e$(%) between the average dispersions and the experimental result, $e = |sim - exp|/exp$ is plotted (in %). This measure quantifies the distance of the dots from the horizontal line in the Top panel. Different JNDs are named after Fig 2. Vertical bars refer to the variation of values in the late stage of the simulation, in the range $1.5 \times 10^6 - 2 \times 10^6$ games per agent.

doi:10.1371/journal.pone.0125019.g003

Overall, smooth JNDs produce a weaker clustering of the colour naming patterns across different populations, and vice versa rough JNDs force a greater uniformity. The amount of universality observed in a set of languages depends therefore (also) on a specific property of the individual bias. Speculatively, this result suggests that, had the human JND been less smooth, we would have observed a greater regularity in colour naming patterns across different languages. It is also important to highlight that the human JND happens to produce a very good agreement with experimental data, and *this is not a trivial finding, since different JNDs can perform much worse*. Said in other words, the model is in fact sensitive to the shape of the JND, and the human JND happens to produce a small discrepancy between the simulations and the WCS data, as shown in Fig 3.

## Departures from universality

In the previous section we focused on the question: how sensitive is the result of our model to the particular form of JND curve? It turned out that the form of the JND function matters, i.e., choices different from the "human" one do not agree, in statistical terms (clustering of languages), with the empirical WCS data. However we cannot underestimate the *dynamical* and *stochastic* nature of our model: indeed, it does not simply optimise some sort of energy landscape associated to the external constraint (e.g. the JND curve). On the contrary, it goes through a sequence of trials, errors, and successes that are not easily forgotten and affect the subsequent evolution of the category system, as pointed out by a recent study based on time-

dependent statistical correlation functions [40]. Such a history-dependent "cultural" process leads to a final state not necessarily optimal with respect to the JND function: for this reason we consider the JND to be a *weak bias* on category formation, as it is not sufficient to determine the category pattern. Let us now address this problem in a quantitative way, focusing on the structure of categories, as represented by their centroids.

Our first aim is to verify a correlation between the average pattern of categories and the JND curve. We simulated 600 independent populations each with $N = 200$ agents. All the simulations were stopped at $10^6$ games per agent, when a quasi-stationary state with high communicative success has been reached. For each population $j \in [1, 600]$, we identified the $m_j$ categories shared by the $N$ agents: their centroids are averaged across the agents, giving rise to the population category pattern $s_j = \{x_1^j, x_2^j, ..., x_{m_j}^j\}$ where $x_i^j \in (0, 1)$ is the average centroid of the $i$-th category of the $j$-th population. Then we collected together in a single set $s$ all the positions $x_i^j$ (with $i \in [1, m_j]$ and $j \in [1, 600]$) of centroids from all categories of all populations. Finally we divided the perceptual space, i.e. the segment [0, 1] into 50 bins and count how many centroid's positions fall in each bin, producing the category histogram. This histogram, normalised in order to give total area 1, is shown in the main graph of Fig 4. We consider such a

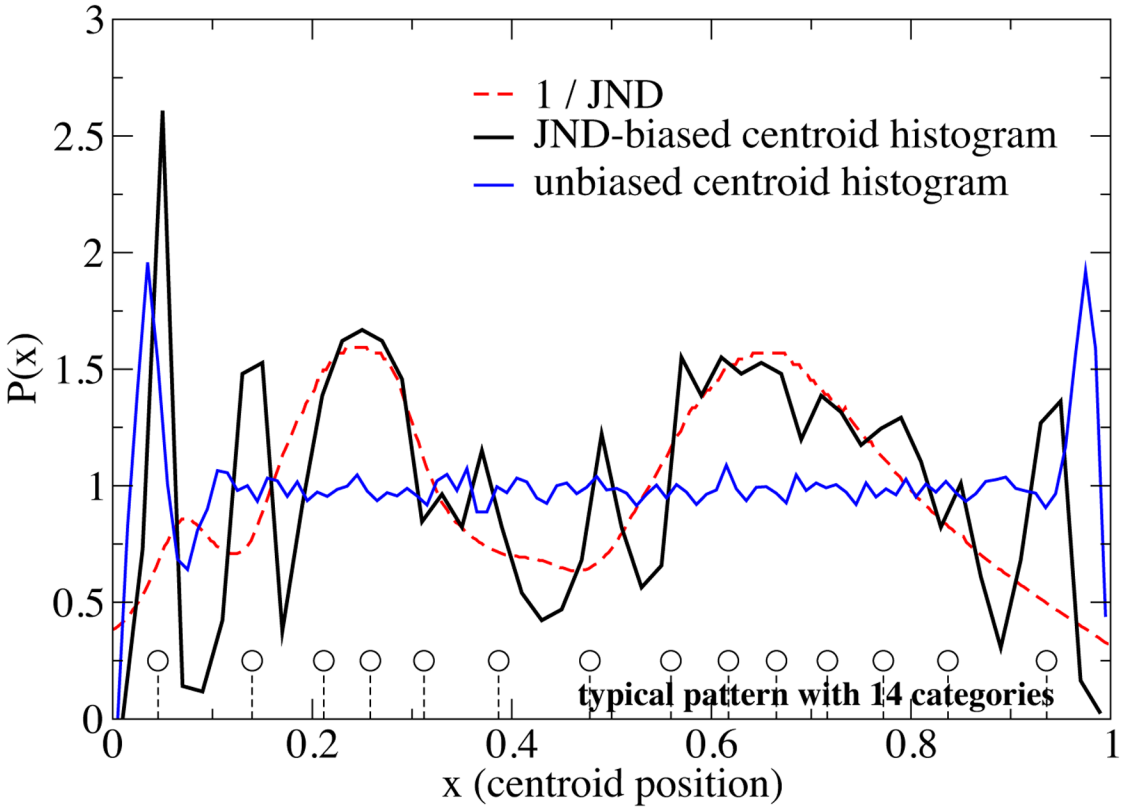


**Fig 4. Numerical World Color Survey with human and neutral JNDs: structure of categories.** The two solid curves represent the histograms of position of category-centroids obtained in simulations with the human (black curve) and neutral JND (blue curve). As a reference, the inverse of the human JND, rescaled by a constant factor, is displayed (red dashed curve). In the central region strong correlation is seen between the centroid distribution from "human-like" simulations and the inverse of the human JND, while the outcome of unbiased simulations is flat in the same region. Strong oscillations near the two extrema (0 and 1) are appreciated in both models, typical of "hard boundaries". The solid circles displayed at the bottom of the graph represent the "average pattern" of the $\sim$ 150 populations which display 14 categories at the end of "human-like" simulations.

doi:10.1371/journal.pone.0125019.g004

histogram, also called *distribution* in the following, as a good estimate of the probability of finding a category centroid in the Category Game informed with a given JND curve. The black curve is the distribution of centroid positions in populations simulated with the human JND, the blue one is obtained with a neutral (flat) JND. The red dashed curve is the inverse of the human JND curve, rescaled by a constant factor (for the purpose of displaying it on the same scale). We observe that—in the central region $\sim$ [0.2, 0.8] of the light spectra—a correlation exists between the probability of finding a centroid and the inverse of the JND curve. Centroids are more easily found where the JND is smaller, i.e., where the perceptual resolution power of agents is higher. Interestingly, large oscillations appear near the two extremes (0 and 1) in both human and neutral simulations. Such oscillations signal the presence of hard boundaries: indeed no centroids can appear too close to 1 or 0 (precisely at distance smaller than the JND in the extrema, which is $\sim$ 0.02 for humans and $\sim$ 0.01 in the neutral model) and therefore an almost periodic pattern is likely to appear near the boundaries. This effect resembles the oscillations of density for liquids of hard spheres near a wall [55]. It is expected that simulations with periodic boundary conditions will not display those oscillations.

Our next step is estimating how large the influence is of the JND on the structure of categories. For this purpose, we restricted our focus to simulations with the human JND. We sampled 5000 different populations. Populations with the human JND tend to have a number of categories between 9 and 19: this number shows a roughly bell-shaped distribution, whose maximum ($\sim$ 1300 populations out of the 5000 simulated) is at $m = 14$. We restricted further our analysis considering only those populations with $m = 14$ categories, in order to single out the largest available group of populations ending up with a given $m$. We computed the average, or "typical", pattern of categories for this sub-group, $\bar{s} = \{\bar{x}_1, \bar{x}_2, ..., \bar{x}_{14}\}$. The 14 average positions of typical centroids are displayed at the bottom of Fig 4.

Having established the *typical pattern*, we went further, asking how far the pattern $s_j$ of the $j$-th population is from the typical one $\bar{s}$. The simplest tool to this end is the Euclidean (squared) distance $d_j = |\, s_j - \bar{s}\, |^2 \equiv \sum_{i=1}^{14} (x_i^{(j)} - \bar{x}_i)^2$ of the $j$-th population from the typical pattern. The $\sim$ 1300 values of $d_j$ obtained are shown in the inset graph of Fig 5. While it is intuitively clear that such a distance can take values distinctly larger than zero, a quantitative assessment of this observation can be obtained by computing the distribution of the values of $d_j$, which we report in the main graph of Fig 5. The distribution appears quite broad: indeed its tail at large values (i.e. $d > 0.01$) decays slower than an exponential. In particular, a possible power-law decay $P(d) \sim d^{-\alpha}$ (with $\alpha \simeq 3$) could approximate such a fat tail.

We stress that a power-law tail is a typical signal of fluctuations *larger than normal*. In the absence of strong correlations, the random wandering of stochastic variables (such as the distance from the average that we are considering) is usually expected to have Gaussian or exponential tails. A simple example is borrowed from the theory of fluctuations in equilibrium statistical physics: a Brownian particle in a harmonic potential displays fluctuations of its distance $x$ from the well's center distributed as $\sim e^{-cx^2}$ where $c$ is a constant depending on the temperature and the curvature of the potential. Einstein's celebrated fluctuation theory [56] generalises this observation and predicts again an exponential distribution $\sim e^S$ with $S$ the system's entropy, which is the quantity maximised at equilibrium (i.e. $-S$ is the optimised cost). Non-exponential distributions typically reveal that fluctuations do not come from a system which tries to optimise a cost function.

Nevertheless, with our available statistics, the "fat" (power-law) nature of our tail cannot be certified. More compelling is the comparison between the results from the Category Game and those from a "Random model" where the 14 centroids are independently and uniformly distributed on the segment [0, 1]. On one hand the red dashed curve in Fig 5, coming from a

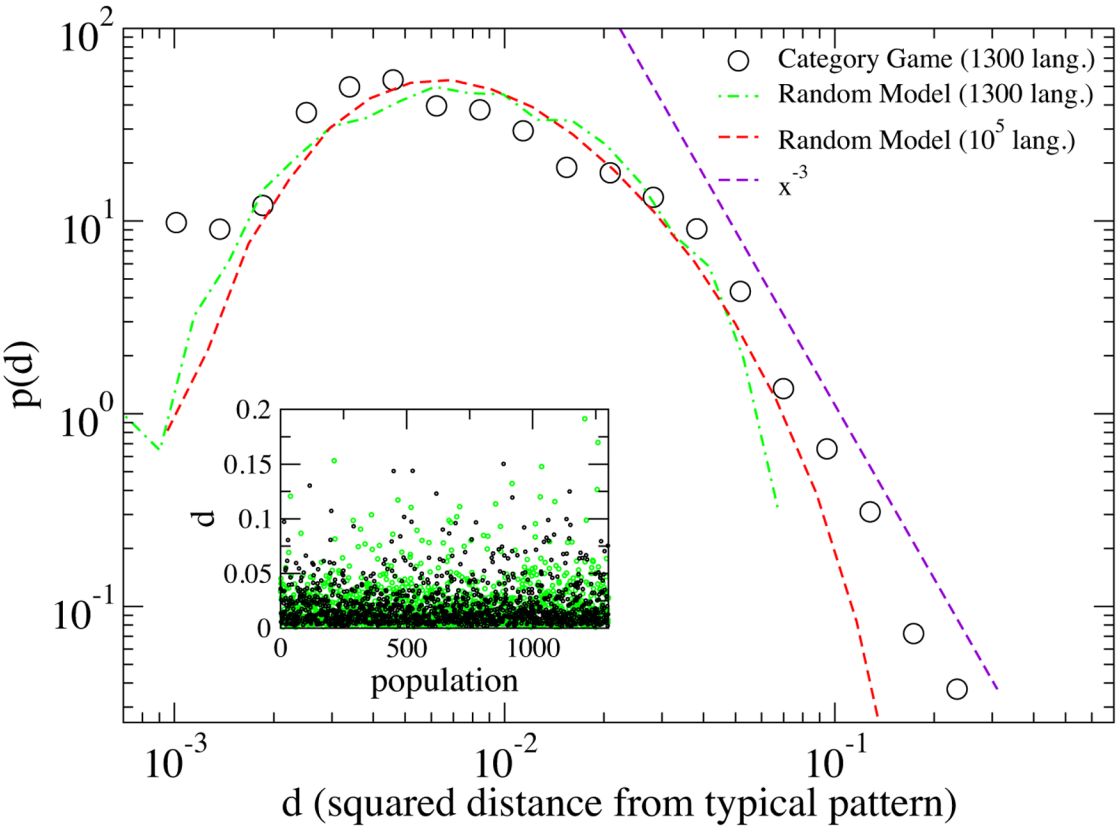


**Fig 5. Numerical World Color Survey with human JNDs: weakness of the bias.** The histogram of the squared distance $d = \sum_{i=1}^{14} (x_i - \bar{x}_i)^2$ between the position of the *i*-th centroid and its average ("typical") value. Black data represent the statistics of those populations displaying 14 linguistic categories at the end of Category Game simulations with the human JND, which were roughly $\sim$ 1300 of the total 5000 considered populations. The inset displays the actual values of *d* for each population. Green and red data come from a "Random" model where each "language" is produced by a uniform random distribution of 14 category centroids. A random case with a very large number of languages (red curve) represents the "ideal" statistics of such a model. The distances *d* from the average pattern (shown in the inset) appear to have a larger average $\bar{d}_{rand}$ than the Category Game model $\bar{d}_{CG}$ (where by $\bar{d}$ we refer to the mean value of the random and the CG cases, respectively), but since we are interested in the fluctuations we have rescaled the random model data, dividing them by $\bar{d}_{rand}/\bar{d}_{CG}$, in order to compare the histograms. A power law fit $\sim x^{-3}$ is also shown as a guide for the eye.

doi:10.1371/journal.pone.0125019.g005

simulation of $10^5$ random languages, illustrates the "ideal" distribution of such a model, which has an exponential cut-off for large distances. On the other hand the green curve allows us to evaluate also finite size effects, by taking into account only 1300 random languages: in such a case the histogram is noisy but close to the ideal one and it is not as broad as the one related to the Category Game.

In conclusion we find that in the Category Game—with a non-negligible probability—a population may display a category pattern quite far from the typical one. Thus, "typical" does not mean "certain". The bias induced by the external constraint (the human JND) is not strong enough to attract all patterns to a typical configuration, and the history-dependent dynamics may lead a population into states only weakly affected by the JND. Repeating such an analysis with different choices of $m \neq 14$ leads to similar results with compatible power-law exponents, signalling the robustness of the result. It might be interesting to repeat the study of these fluctuations of "distances" from the typical pattern with the actual data of the WCS. However the number of available languages is too small, and restricting the analysis to a given number of

colour terms would reduce the statistics even more. For these reasons we did not pursue this idea.

## Discussion

We have analysed the mechanism through which a weak cognitive or perceptual bias ends up influencing the structure of a language, focusing on the prototypical case of colour categorisation and studying the Category Game model. Previous studies showed that in this framework the presence of the JND bias triggers the emergence of universal colour naming patterns whose statistical properties are in agreement with the ones observed in the WCS [30], and that the model is able to reproduce the observed hierarchy of colour names [31]. Here we have seen that:

1. The particular form assumed by the perceptual bias (i.e., the shape of the JND function) does affect the properties of the universality patterns generated by the model. The fact that the human JND is responsible for a quantitative agreement with the WCS data (see discussion above and [30], [31]) is therefore genuinely significant.
2. A further source of biases arises from the dynamical evolution of the system. While the JND is universal and would tend to result in highly similar category patterns in different languages, the specific evolution of each language acts in the opposite direction by introducing frozen accidents and randomness in the process. Quantitatively, the distribution of the distances from the typical pattern of the various languages exhibits a fat tail, which indicates the possibility of significant, culture-dependent, deviations. Our study of fluctuations suggests that colour categorisation is not a simple optimisation process where some cost function is minimised.

These conclusions arise from a series of numerical experiments in which the agents were endowed with artificially manipulated JNDs. In this respect, a remark is in order on the agreement with the data obtained when the human JND is considered (at least as far as the test introduced in [15] and the hierarchy of colour names are considered). Whether this is an (extremely) fortunate coincidence or has more profound reasons can not be "proved" in this context. The task of simulations is that of testing theoretical hypotheses, showing whether they are realistic, and checking for the implications [49].

What we have shown is that a perceptual bias *can* in fact induce universality, to be intended in a statistical sense, and that the shape of this bias can influence the degree of universality (i.e., regularity across different languages). The quite large cross-linguistic variability of the colour naming patterns emergent in population endowed with the human JND indicates that strong deviations from the most common patterns are possible. The broadness of the fluctuations distribution suggests that its origin could be dynamical, i.e., related to a single population's history. This clarifies why universality has to be intended in a statistical sense, and that the history of a language can reduce the tendency of a shared bias to enforce uniformity.

## Appendix: The Category Game model

The game is played by a population of $N$ individuals. Each individual is characterised by its partition of the [0, 1) perceptual channel in non-overlapping contiguous segments henceforth called *(perceptual) categories*. Each category has an associated inventory of words, that constitutes its linguistic counterpart. Individuals are endowed with the capability of transmitting words to each other and to interact non-linguistically through pointing at objects in the environment.

At each time step $t = 1, 2, ..$ two individuals are randomly extracted and interact, one playing as speaker and the other one as hearer. They face a *scene* of *M objects*, i.e. *M* real numbers randomly extracted from the interval [0, 1), with $M \geq 2$. One of the objects is the topic *h* that the speaker will try to communicate to the hearer.

The interaction involves the following steps:

## Discrimination

The speaker perceives the scene, i.e. assigns each object $i \in [0, 1)$ of the scene to one of its categories. The category associated with object *i* is the unique segment $[l, r)$ of the individual perceptual channel for which it holds $l \leq i < r$. An object *k* is said to be *discriminated* by category *C* if *k* is the only object of the scene to be associated to *C*. In other words, if *k* is discriminated by *C*, then given any object *j* in the scene it holds $j \in C \Leftrightarrow j \equiv k$.

There are two possibilities for the topic *h*:

**Either** *h* is already discriminated by a category *C*;

**or** *h* and a non empty set *O* of different objects *i* fall in the same category *C*.

In the latter case the speaker refines the category partition of its perceptual channel to discriminate the topic. Given the two objects *a*, *b* for which it holds $a \equiv \max_{i \in O}\{i{:}i < h\}$ and $b \equiv \min_{i \in O}\{i{:}i > h\}$, category *C* is split in new categories by the introduction of new boundaries in $(a+h)/2$ and $(h+b)/2$. [If $h > i, \forall i$ $(h < i, \forall i)$ then *a* (*b*) is not defined, and of course the corresponding new boundary is not created.] Each new category inherits the linguistic inventory of *C*, plus a brand new word.

## Word transmission

After discrimination, the speaker transmits a word to the hearer in order to identify the topic.

**If** a previous successful communication event has occurred with the discriminating category the speaker transmits the word that yielded that success;

**else** the speaker transmits the brand new word added to the discriminating category when it was created.

## Word reception

The hearer receives the transmitted word, and, looking at its repertoire, identifies the set of all categories

1. whose inventories contain the transmitted word and
2. that are associated to at least one object in the scene.

## Guessing and outcomes of the game

There are now several mutually exclusive possibilities for the hearer:

- **a** The set is empty;
- **b** The set contains only one category, corresponding to a single object in the scene;
- **c** The set contains only one category, corresponding to more than one object in the scene;
- **d** The set contains more than one category.

Then:

- **if a** The hearer cannot infer which is the topic, and communicates its perplexity to the speaker (we can image that individuals have a built in conventionalised way of doing that, for instance pointing at the sky).
- **if b** There is only a candidate object for the hearer, who points at it;
- **if c or d** The hearer points randomly at one of its candidate objects.

At this point the speaker unveils the topic (pointing at it), and both individuals become aware of the result of their interaction, that is

**success** if the object pointed by the hearer corresponds to the topic or
**failure** in all the other cases.

### Updating

Independently of the outcome of the game, the hearer checks whether the topic is discriminated by one of its categories. If this is not the case, it discriminates the topic following the discrimination procedure described above.

Then:

- **in case of failure** the hearer adds the transmitted word to the category discriminating the topic;
- **in case of success** both agents delete all the words but the transmitted one from the inventory of the category discriminating the topic.

## Author Contributions

Conceived and designed the experiments: AB VL AP. Performed the experiments: AB VL AP. Analyzed the data: AB VL AP. Wrote the paper: AB VL AP.

## References


1. Croft W. Typology and universals. Cambridge Univ Pr; 2002.
2. Bickerton D, Szathmáry E. Biological Foundations and Origin of Syntax. The MIT Press; 2009.
3. Smith ADM, De Boer B, Schouwstra M. The Evolution of Language: Proceedings of the 8th International Conference (EVOLANG8). World Scientific Publishing Company; 2010.
4. Chomsky N. Rules and representations. Blackwell, Oxford, UK; 1995.
5. Pinker S. The language instinct. Harper Collins, New York, USA; 1995.
6. Christiansen MH, Chater N. Language as shaped by the brain. Behavioral and Brain Sciences. 2008; 31(05):489–509. doi: 10.1017/S0140525X08004998 PMID: 18826669
7. Chater N, Reali F, Christiansen MH. Restrictions on biological adaptation in language evolution. Proceedings of the National Academy of Sciences. 2009; 106(4):1015. doi: 10.1073/pnas.0807191106
8. Kirby S, Dowman M, Griffiths TL. Innateness and culture in the evolution of language. Proceedings of the National Academy of Sciences. 2007; 104(12):5241. doi: 10.1073/pnas.0608222104
9. Griffiths TL, Kalish ML. Language evolution by iterated learning with Bayesian agents. Cognitive Science. 2007; 31(3):441–480. doi: 10.1080/15326900701326576 PMID: 21635304
10. Evans N, Levinson SC. The myth of language universals: Language diversity and its importance for cognitive science. Behavioral and Brain Sciences. 2009; 32(05):429–448. doi: 10.1017/S0140525X0999094X PMID: 19857320
11. Deacon TW. The symbolic species: The co-evolution of language and the brain. New York, USA: Norton & Company; 1998.
12. Baronchelli A, Chater N, Pastor-Satorras R, Christiansen MH. The biological origin of linguistic diversity. PloS one. 2012; 7(10):e48029. doi: 10.1371/journal.pone.0048029 PMID: 23118922

13. Berlin B, Kay P. Basic Color Terms. Berkeley: University of California Press; 1969.
14. Cook RS, Kay P, Regier T. The World Color Survey database: history and use. Handbook of Categorisation in the Cognitive Sciences Amsterdam and London: Elsevier. 2005;. doi: 10.1016/B978-008044612-7/50064-0
15. Kay P, Regier T. Resolving the question of color naming universals. Proc Natl Acad Sci USA. 2003; 100(15):9085–9089. doi: 10.1073/pnas.1532837100 PMID: 12855768
16. Taylor JR. Linguistic categorization. Oxford University Press New York; 2003.
17. Witkowski SR, Brown CH. An Explanation of Color Nomenclature Universale. American Anthropologist. 1977; 79(1):50–57. doi: 10.1525/aa.1977.79.1.02a00050
18. Comrie B. Language universals and linguistic typology: Syntax and morphology. University of Chicago press; 1989.
19. Gardner H. The Mind's New Science: A History of the Cognitive Revolution. New York: Basic Books; 1985.
20. Lakoff G. Women, fire, and dangerous things: What categories reveal about the mind. Chicago: University of Chicago Press; 1987.
21. Murphy GL. The big book of concepts. Bradford Book; 2004.
22. Steels L, Belpaeme T. Coordinating perceptually grounded categories through language: A case study for colour. Behav Brain Sci. 2005; 28(04):469–489. doi: 10.1017/S0140525X05000087 PMID: 16209771
23. Belpaeme T, Bleys J. Explaining universal color categories through a constrained acquisition process. Adap Behv. 2005; 13(4):293–310. doi: 10.1177/105971230501300404
24. Dowman M. Explaining Color Term Typology With an Evolutionary Model. Cog Sci. 2007; 31(1):99–132. doi: 10.1080/03640210709336986
25. Komarova NL, Jameson KA, Narens L. Evolutionary models of color categorization based on discrimination. J Math Psych. 2007; 51(6):359–382. doi: 10.1016/j.jmp.2007.06.001
26. Komarova NL, Jameson KA. Population heterogeneity and color stimulus heterogeneity in agent-based color categorization. Journal of Theoretical Biology. 2008; 253(4):680–700. doi: 10.1016/j.jtbi.2008.03.030 PMID: 18534626
27. Puglisi A, Baronchelli A, Loreto V. Cultural route to the emergence of linguistic categories. Proc Natl Acad Sci USA. 2008; 105(23):7936. doi: 10.1073/pnas.0802485105 PMID: 18523014
28. Jameson KA, Komarova NL. Evolutionary models of color categorization. I. Population categorization systems based on normal and dichromat observers. JOSA A. 2009; 26(6):1414–1423. doi: 10.1364/JOSAA.26.001414 PMID: 19488181
29. Jameson KA, Komarova NL. Evolutionary models of color categorization. II. Realistic observer models and population heterogeneity. JOSA A. 2009; 26(6):1424–1436. doi: 10.1364/JOSAA.26.001424 PMID: 19488182
30. Baronchelli A, Gong T, Puglisi A, Loreto V. Modeling the emergence of universality in color naming patterns. Proceedings of the National Academy of Sciences. 2010; 107(6):2403. doi: 10.1073/pnas.0908533107
31. Loreto V, Mukherjee A, Tria F. On the origin of the hierarchy of color names. Proceedings of the National Academy of Sciences. 2012; 109(18):6819–6824. doi: 10.1073/pnas.1113347109
32. Narens L, Jameson KA, Komarova NL, Tauber S. Language, categorization, and convention. Advances in Complex Systems. 2012; 15(03n04). doi: 10.1142/S0219525911500226
33. Endrikhovski S. A computational model of color categorization based on statistics of natural images. Journal of Vision. 2002; 2(10):16–16. doi: 10.1167/2.10.16
34. Xu J, Dowman M, Griffiths TL. Cultural transmission results in convergence towards colour term universals. Proceedings of the Royal Society B: Biological Sciences. 2013; 280 (1758). doi: 10.1098/rspb.2012.3073
35. Wittgenstein L. Philosophical Investigations. (Translated by Anscombe G.E.M.). Oxford, UK: Basil Blackwell; 1953.
36. Steels L. A self-organizing spatial vocabulary. Artificial Life. 1995; 2(3):319–332. doi: 10.1162/artl.1995.2.3.319 PMID: 8925502
37. Bedford R, Wyszecki G. Wavelength discrimination for point sources. JOSA. 1958; 48(2):129–130. doi: 10.1364/JOSA.48.000129
38. Long F, Yang Z, Purves D. Spectral statistics in natural scenes predict hue, saturation, and brightness. Proc Natl Acad Sci USA. 2006; 103(15):6013–6018. doi: 10.1073/pnas.0600890103 PMID: 16595630

39. Dunn M, Greenhill SJ, Levinson SC, Gray RD. Evolved structure of language shows lineage-specific trends in word-order universals. Nature. 2011; 473(7345):79–82. doi: 10.1038/nature09923 PMID: 21490599

40. Mukherjee A, Tria F, Baronchelli A, Puglisi A, Loreto V. Aging in language dynamics. PloS ONE. 2011; 6(2):e16677. doi: 10.1371/journal.pone.0016677 PMID: 21390207

41. Boyd R, Richerson PJ. Culture and the evolutionary process. University of Chicago Press; 1988.

42. Kirby S. Spontaneous evolution of linguistic structure-an iterated learning model of the emergence of regularity and irregularity. Evolutionary Computation, IEEE Transactions on. 2001; 5(2):102–110. doi: 10.1109/4235.918430

43. Bayes T. An Essay towards solving a Problem in the Doctrine of Chances. Philosophical Transactions (1683–1775). 1763;p. 370–418.

44. Regier T, Kay P, Khetarpal N. Color naming reflects optimal partitions of color space. Proceedings of the National Academy of Sciences. 2007; 104(4):1436. doi: 10.1073/pnas.0610341104

45. Jameson K, DAndrade RG. 14 It's not really red, green, yellow, blue: an inquiry into perceptual color space. Color categories in thought and language. 1997;p. 295.

46. Kay P. Asymmetries in the distribution of composite and derived basic color categories. Behavioral and Brain Sciences. 1999; 22(06):957–958. doi: 10.1017/S0140525X99372216

47. Xiao Y, Kavanau C, Bertin L, Kaplan E. The Biological Basis of a Universal Constraint on Color Naming: Cone Contrasts and the Two-Way Categorization of Colors. PloS one. 2011; 6(9):e24994. doi: 10.1371/journal.pone.0024994 PMID: 21977226

48. De Beule J, Bleys J. Self organization and emergence in language: a case study for color. In: The evolution of language: proceedings of the 8th International Conference (EVOLANG8), Utrecht, Netherlands, 14–17 April 2010. World Scientific Pub Co Inc; 2010. p. 83.

49. Jaeger H, Baronchelli A, Briscoe T, Christiansen MH, Griffiths T, Jäger G, et al. What can mathematical, computational and robotic models tell us about the origins of syntax? In: Bickerton D, Szathamary E, editors. Biological Foundations and Origin of Syntax. Strungmann Forum Reports, vol. 3. Cambridge, MA: MIT Press; 2009.

50. Baronchelli A, Felici M, Caglioti E, Loreto V, Steels L. Sharp Transition towards Shared Vocabularies in Multi-Agent Systems. Journal of Statistical Mechanics. 2006; P06014.

51. Bray AJ. Theory of Phase Ordering Kinetics. Advances in Physics. 1994; 43:357. Available from: http://www.citebase.org/abstract?id = oai:arXiv.org:cond-mat/9501089. doi: 10.1080/00018739400101505

52. Cavagna A. Supercooled liquids for pedestrians. Physics Reports. 2009; 476(4–6):51–124. doi: 10.1016/j.physrep.2009.03.003

53. Kay P, Berlin B, Merrifield W. Biocultural implications of systems of color naming. Journal of Linguistic Anthropology. 1991; 1(1):12–25. doi: 10.1525/jlin.1991.1.1.12

54. Plümacher M, Holz P. Speaking of colors and odors. vol. 8. John Benjamins Publishing; 2007.

55. Hansen JP, McDonald IR. Theory of Simple Liquids. Academic Press; 2006.

56. Einstein A. Theorie der Opaleszenz von homogenen Flüssigkeiten und Flüssigkeitsgemischen in der Nähe des kritischen Zustandes. Annalen der Physik. 1910; 338(16):1275–1298. doi: 10.1002/andp.19103381612