# Supporting Information

## Loreto et al. 10.1073/pnas.1113347109

### Experimental Hierarchy of Color Names

Berlin and Kay's classic study (1) on typological properties of color vocabularies established the universal presence of a special subset of color names which they called the "basic color names." These are the most salient and frequently used color words across the majority of the world's languages. They represent the following 11 English color names: black, white, red, green, yellow, blue, brown, orange, purple, pink, and gray. Berlin and Kay found that these names have prototype properties, which means that there is usually one name that best represents a color, whereas other colors that are progressively more dissimilar with this color become less good examples for the name. They also found that the number of basic color names range from 2 to 11 across the world's languages, of course with exceptions like Russian and Hungarian which have 12 basic names. A third and a totally unexpected finding by them is that if a language encodes fewer than 11 names, then there are strict limitations on which names it may encode. The typological regularities observed by them can be summarized by the following implicational hierarchy

$$\begin{bmatrix}\text{white}\\ \text{black}\end{bmatrix} < [\text{red}] < \begin{bmatrix}\text{green}\\ \text{yellow}\end{bmatrix} < [\text{blue}] < [\text{brown}] < \begin{bmatrix}\text{purple}\\ \text{pink}\\ \text{orange}\\ \text{gray}\end{bmatrix},$$

where for distinct color names $a$ and $b$, the expression $a < b$ signifies that $a$ is present in every language where $b$ is present but not vice versa. Based on the above observation, the authors further theorize that as languages evolve they acquire the new basic color names in a fixed chronological sequence of the form

Stage I: dark-cool and light-warm.
Stage II: red (including all shades of violet).
Stage III: either green or yellow.
Stage IV: both green and yellow.
Stage V: blue.
Stage VI: brown.
Stage VII: purple, pink, orange, or gray.

We stress how stage I of ref. 1 is not referring to the emergence of the two achromatic colors "black" and "white," rather it refers to a division of the perceptual space that has nothing to do with the chromatic properties of light, being based exclusively on the light intensity. Ratliff writes (2) that the well-known studies of Dani color terms by Heider-Rosch and Olivier (3) "put the question of psychophysiological bases of the two color terms of stage I into better perspective. These terms appear to be panchromatic, more or less equivalent to the general panchromatic English terms dark and light or dull and brilliant rather than equivalent to the specific achromatic terms black and white. Although the Dani color terms do include chromatic colors, and do have attributes of coolness and warmth, the division between them appears to be based mainly on brightness."

It is also important to mention here that six languages studied by Berlin and Kay do not conform to the above presented hierarchy. In some cases this deviation is because there is no basic color name that can be consistently identified with certain parts of the visible spectrum. For instance, as has been noted by Dowman (4), the Kuku-Yalanji (Australia) language has no consistent name for green. Whereas some speakers identify either just green or both green and blue with *kayal*, most of them do not use this name at all for green. Moreover, it should be noted that certain other languages studied by Berlin and Kay appear in a transition between the evolutionary stages because some speakers (especially younger speakers) are found to use more color names than the others (see ref. 4 and the references therein). In the Category Game (CG) framework that we propose here, these deviations from the hierarchy can be naturally attributed to the inherent stochasticity of the underlying cultural dynamics and the fluctuations in the fixation times of the color names in the form of error bars (see Fig. 4 *C* and *D* of the main text) confirm this picture. At the same time, this framework is able to spell out the major characteristics of this hierarchy (in terms of the mean fixation times of the color names) in a remarkable way. In summary, both the universal trends in color naming as well as the possible exceptions to this universality is explained in this framework.

**The Category Game model.** The basic purpose of the CG model (5) is to examine how a population of interacting individuals can develop, through a series of language games, a shared form-meaning repertoire from scratch and without any preexisting categorization. The model involves a set of $N$ artificial agents committed to the task of categorizing a single analogical perceptual channel (e.g., the hue dimension of the color spectrum), each stimulus being represented as a real-valued number ranging in the interval $[0, 1)$. We identify categorization as a partition of the $[0, 1)$ interval (representing the perceptual channel of the agents) into discrete subintervals which are denoted as perceptual categories. Each individual has a dynamical inventory of form-meaning associations linking perceptual categories (meanings) to words (forms), denoting their linguistic counterpart. The perceptual categories as well as the words associated to them coevolve dynamically through a sequence of elementary communication interactions, usually referred to as games. All the players are initialized with only the trivial $[0, 1)$ perceptual category that has no name associated to it. In each step, a pair of individuals (one playing as speaker and the other as hearer) is randomly selected from the population and presented with a new "scene"—i.e., a set of $M \geq 2$ objects (stimuli) where each object is a real number in the $[0, 1)$ interval. (For simplicity and without any loss of generality we assume $M = 2$.) The speaker discriminates the scene and names one object (i.e., the topic) and the hearer tries to guess the topic from the name. A correct guess results in a successful communication. Based on the outcomes of the game, the two individuals update their category boundaries and the inventory of the associated words. A detailed description of the game is provided in Fig. S1.

The perceptive resolution power of the individuals limits their ability to distinguish between the objects in the scene that are too close to each other in the perceptual space. In order to take this factor into account, 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, usually termed the Just Noticeable Difference (JND), encodes the finite resolution power of human vision by virtue of which the artificial agents are not required to distinguish between those categories that a human eye cannot differentiate (see Fig. 1 in the main text).

**Dynamical Properties of the Multilevel Emergence.** ***Evolution of the category structure.*** In the CG dynamics, one can identify two different phases. In the first phase, the number of perceptual categories increases due to the pressure of discrimination, and at the same time many different words are used by different agents for naming similar perceptual categories. This kind of synonymy is found to reach a peak and then suddenly drop, as shown in refs. 5

and 6. Subsequently, a second phase begins when most of the perceptual categories are associated with only one word (6). At this point, words are found to expand their dominion across adjacent perceptual categories. Therefore, sets of contiguous perceptual categories sharing the same words are formed at the different existing levels, giving raise to a single linguistic category (Fig. S2 *A* and *B*). Consequently, an important outcome is the emergence of a hierarchical category structure made of a basic layer, responsible for fine discrimination of the environment, and shared linguistic layers that groups together perceptions at the different levels to guarantee communicative success. Remarkably, the number of linguistic categories in the second phase turns out to be finite and small for all the different levels with a very high agreement among the agents in the population (Fig. S2 *C* and *D*).

***Dependence of the levels on $d_{min}$.*** In the multilevel CG, the emergence of a higher level is strongly tied to the value of $d_{\min}$ chosen. If $d_{\min}$ is high (see Fig. S3*A*), then a third level (level 2) never emerges as a separate entity; in contrast, it mimics the lower level. However, if $d_{\min}$ is low, then a third level is also found to emerge (see Fig. S3*B*) although still in its transient phase after a billion games per player. This observation can be intuitively explained: When $d_{\min}$ is low, two objects at the same distance in the $[0, 1)$ interval more likely belong to different perceptual categories. Because the number of linguistic categories does not depend on $d_{\min}$ (which is one of the main results of the CG), the probability of having a "failure with name" and hence of the emergence of multiple levels increases when decreasing $d_{\min}$. Low $d_{\min}$ allows then for a much "fine-grained" categorization of the perceptual space, which is typically the case with specialized linguistic communities (e.g., painters). On the other hand, the timescales for the emergence of linguistic categories, at any level of categorization, increase while decreasing $d_{\min}$, which explains why regions of the color spectrum corresponding to high $d_{\min}$ are the first to be named with high consensus in the population.

***Fraction of games in the higher level.*** Here we measure the fraction of games that are being played at level 1 when JND is set to $\bar{d}_{\min}$ as well as $d_{\min}(x)$ (Fig. S4*A*). Note that the fraction of games being played in the higher level is proportional to the value of JND chosen [$d_{\min}(x)$ can take up much lower values than its average value $\bar{d}_{\min}$]. This result is simply an outcome of the fact that in case of low values of $d_{\min}$ the number of choices for the topic and the object is much larger, which in turn increases the chances of failure with name eventually resulting in more games being played in the higher level.

Fig. S4*B* illustrates the number of games played over time sliding windows in level 1 in the seven individual regions (see Fig. 1 of the main text) expressed as a fraction of the total number of games played in these regions in level 1 over the same time window. Once again a clear ordering emerges at the onset of the dynamics which is in agreement with the results presented in the main text: This fraction is least in the regions corresponding to low $d_{\min}$ (i.e., regions 4 and 6).

***Extension.*** We define extension as the portion of the $[0, 1)$ space that already has at least one name in a particular level. Note that by definition the extension of level 0 is always 1. Fig. S5 *A* and *B* reports the average extension in the population versus $t/N$ when JND is set to $\bar{d}_{\min}$ and $d_{\min}(x)$, respectively. Both the results indicate that the level 1 is already completely created filling the entire $[0, 1)$ space as soon as $t/N > 10^4$. However, in case where a third level emerges, it only shows up after roughly a million games per player.

***Regional agreement.*** Here we present additional results indicating the emergence of the regional agreement besides that already reported in the main text (Fig. 4 of the main text). In Fig. S6*A* we plot the average match in level 0 for the seven different regions, this time the length of the region being $[c_i - \bar{d}_{\min}, c_i + \bar{d}_{\min}]$: The length of all seven regions in this case is the same and therefore independent of $d_{\min}(x)$ (unlike Fig. 4 of the main text). One observes even for fixed-length regions a time ordering of the emergent agreement that is fully consistent with the result presented in Fig. 4 of the main text. Therefore, it is reasonable to conclude that this effect is independent of the length of the regions chosen and is completely determined by the centers (and the corresponding $d_{\min}$) of these regions.

Fig. S6*B* illustrates how the success rate emerges in these seven fixed-length regions. Note that if the agents are successful in any of the levels, then the outcome of the game is assumed to be a success. Success rate is the fraction of successful games over time sliding windows. This quantity is an alternative measure of the agreement among the agents (more successful games result from a larger agreement) and reflects a very similar time ordering as observed in Fig. S6*A*. Fig. S6*C* shows how the success rate emerges in the seven variable-length regions defined in the main text for Fig. 4. Once again, a similar time ordering discussed in all the previous results is observed.

Finally, in Fig. S6*D*, we plot the emergent match at level 1 in the seven variable-length regions. Although a high agreement is reached for all the regions, no clear time ordering that could be correlated to the corresponding $d_{\min}$ is found to emerge, implying that it is hard to arrange complex color names in a clear hierarchy, unlike the basic color names.

**Control Experiments.** In this section, we show that the similarity of the color order obtained from our results to that from the World Color Survey (WCS) in not a pure coincidence. As a final check for the robustness of our results we perform the following control experiments. We consider, in particular, two null situations where we endow the agents with: (*i*) a flat JND (i.e., $\bar{d}_{\min} = 0.0143$), which is the average value of the human JND (as it is projected on the $[0, 1)$ interval) and (*ii*) the (properly rescaled) inverse of the JND function (Fig. S7).

In both cases, the outcomes of the dynamical evolution have to be compared with that obtained using the actual human JND function [i.e., $d_{\min}(x)$]. Fig. S8 *A* and *B* reports the emergence of agreement (in terms of match) and the fixing times, respectively, for the case of flat JND.

Fig. S8 *C* and *D* reports the emergence of agreement and the fixing times, respectively, for the case of inverse JND. Importantly, here the fixing times are plotted against the wavelength values of the light corresponding to the seven regions considered in the main text. Note that, in none of these two cases the hierarchy obtained from the WCS is reproduced. In the case of flat JND, the fixing times for all the seven regions are roughly equal, whereas in the case of inverse JND, the fixing times are nearly opposite to what we find in case of the actual JND. Thus, the outcome reported in the manuscript is not a pure coincidence and only a right choice of JND coupled on top of it with a complex dynamical process of nonlinear interactions can reproduce the color hierarchy observed across human languages.

**Effects of Rotation of the Stimuli.** In order to further establish the robustness of our results, we repeat our experiments, however, with the stimuli now rotated. In particular, the topic values (i.e., $x$) are given a shift such that $x \leftarrow (x + 0.5) \bmod 1$ (Fig. S9 *A*). Note that this shift brings the regions corresponding to "red" and "violet" at the center and therefore close to each other, while placing regions corresponding to "orange" and "green" at the two distant ends. For this experiment, the agents are endowed with this rotated form of the JND function. Fig. S9*B* shows how the agreement (i.e., match) emerges in the seven variable-length regions defined in the main text for Fig. 4, however, now with the

centers appropriately shifted [i.e., $c \leftarrow (c + 0.5) \bmod 1$]. The results clearly indicate that, even under rotation, the order in which the agreements emerge in the seven regions remain precisely similar to that reported in Fig. 4 of the main text. Therefore, it is reasonable to advocate that the color order obtained from our results is not a pure coincidence; rather it shows how a culturally driven phenomenon can actually lead to the emergence of the color hierarchy that is strongly correlated to the empirical observations so far made in the literature.

**Toward an Evolutionary Dynamics.** In order to further check the robustness of our results, we ran an additional experiment along the following lines. We imagine that each individual performs a certain number of linguistic interactions along her own lifetime. Instead of considering the lifetime infinite, we introduced a rate of replacement of each individual—i.e., the rate over which a given individual is removed from the population and substituted by a "blank slate" individual. This process mimics the birth of new individuals (e.g., children) who have to learn the current language. The new individual could in principle affect the current language (i.e., the shared categorization that emerged so far) and the main question is whether the population is still able to bootstrap a shared system of linguistic categories. We ran the experiment using several values of the replacement rate and the results are reported in Fig. S10.

It is evident how the agreement is always reached and the replacement rate introduces only a small perturbation to the overall dynamics. The replacement rate $r$ is directly connected to the lifetime duration. It takes $N/r$ steps to replace on average the whole population. This value corresponds to $1/r$ games per agent and this number estimates the duration of a generation as well as the average lifetime of an individual. In our case, we are ranging from $10^3$ to $10^5$ games per agent. Notice that for the time being we are not yet simulating a truly genetic algorithm because there is no notion of fitness in there. A thorough investigation of the interplay between cultural and evolutionary timescales will also be presented elsewhere to compare our results with the outcomes of the approaches based on the so-called Iterated Learning Model (7).


1. Berlin B, Kay P (1969) *Basic Color Terms* (Univ California Press, Berkeley, CA).
2. Ratliff, F (1976) On the psychophysiological bases of universal color terms. *Proc Am Philos Soc* 120(5):311–330.
3. Heider-Rosch E, Olivier DC (1972) The structure of the color space in naming and memory for two languages. *Cogn Psychol* 3:337–354.
4. Dowman M (2007) Explaining color term typology with an evolutionary model. *Cogn Sci* 31:99–132.
5. Puglisi A, Baronchelli A, Loreto V (2008) Cultural route to the emergence of linguistic categories. *Proc Natl Acad Sci USA* 105:7936.
6. Mukherjee A, Tria F, Baronchelli A, Puglisi A, Loreto V (2011) Aging in language dynamics. *PLoS One* 6:e16677.
7. Smith K, Kirby S, Brighton H (2003) Iterated learning: A framework for the emergence of language. *Artif Life* 9:371–386.


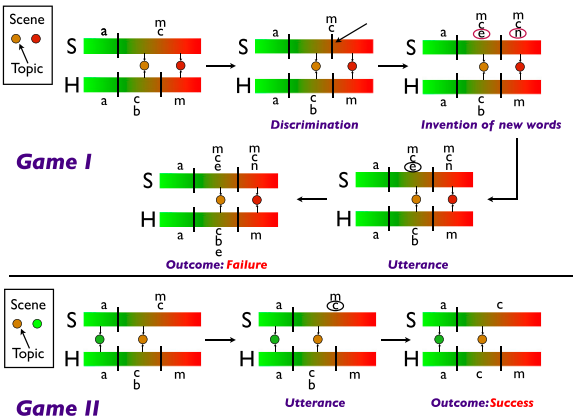


**Fig. S1.** The basic steps of the CG. A pair of examples representing a failure (game I) and a success (game II), respectively. In a game, two players (S denoting the speaker and H denoting the hearer) are randomly selected from the population. Both the players are presented with a scene with two objects and the speaker selects the topic for the subsequent communication. In game I, because the two objects belong to the same perceptual category of the speaker, the speaker has to discriminate her perceptual space by creating a boundary at the middle of the segment containing the two objects (marked by the bold black arrow). The two new categories formed after discrimination inherit the words inventory of the parent perceptual category (here the words "m" and "c"); in addition, a different brand new word is invented for each of the two categories (words "e" and "n" marked by colored circles). Subsequently, the speaker browses the list of words associated to the perceptual category containing the topic (i.e., m, c, and e here). At this point, there can be two possibilities: If a previous successful communication has occurred with this category, the last winning word is chosen; alternatively, the last word invented is selected. For the current example, the speaker chooses the word e (marked by the black circle here), and transmits it to the hearer. The outcome of the game is a failure because the hearer does not have the word e in her inventory associated with the topic. Finally, the speaker unveils the topic, in a nonlinguistic way (e.g., by pointing at it), and the hearer adds the new word to the word inventory of the category corresponding to the topic. In game II, the topic that the speaker chooses is already discriminated. Therefore, the speaker verbalizes it using the word c (which, for example, is possibly the winning word in the last successful communication concerning that category). The hearer knows this word and can therefore point to the topic correctly, thereby leading to a successful game. Both the players dispose all competing words for the perceptual category corresponding to the topic except c. In general, if there are ambiguities (e.g., the hearer finds the word uttered to be linked to multiple categories containing an object), they are resolved by making an unbiased random choice of one of the categories.

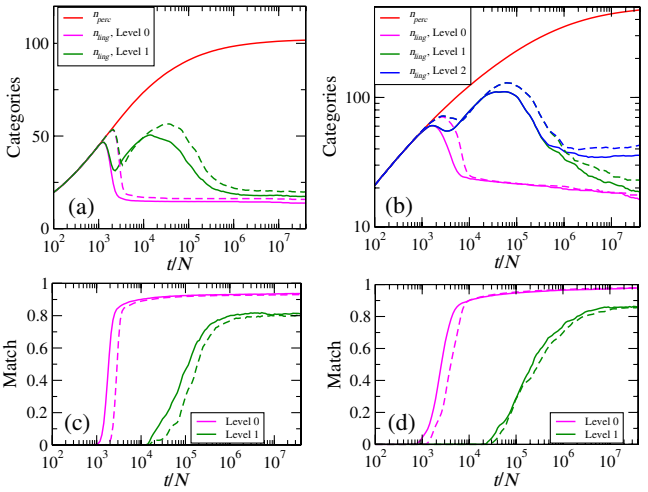


**Fig. S2.** Evolution of the category structure. (*A*) Evolution of the average number of perceptual categories as well as the average number of linguistic categories at different levels when JND is set to $\bar{d}_{\min}$. (*B*) Evolution of the average number of perceptual categories as well as the average number of linguistic categories at different levels when JND is set to $d_{\min}(x)$. (*C*) The average match in the population at different levels versus $t/N$ when JND is set to $\bar{d}_{\min}$. (*D*) The average match in the population at different levels versus $t/N$ when JND is set to $d_{\min}(x)$. Solid lines show results for $N = 300$ and broken lines show results for $N = 700$. All the results are averaged over 30 simulation runs.

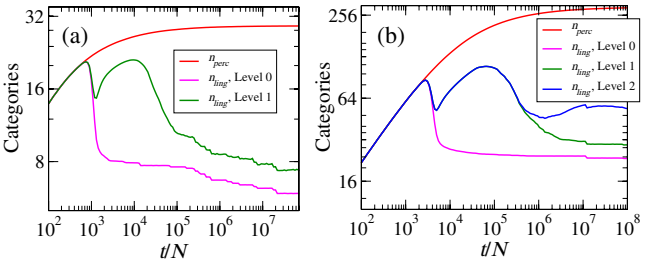


**Fig. S3.** Dependence of the higher levels on $d_{\min}$. Evolution of the average number of perceptual categories as well as the average number of linguistic categories at different levels when JND is set to (*A*) $d_{\min} = 0.05$ and (*B*) $d_{\min} = 0.005$. The results are shown for $N = 500$ and are averaged over 30 simulation runs.

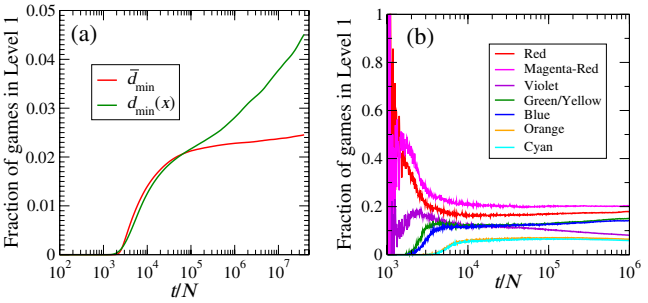


**Fig. S4.** The fraction of games played by the agents in level 1. (*A*) The fraction of the total number of games being played in level 1 versus $t/N$ when the value of JND is set to $\bar{d}_{\min}$ as well as $d_{\min}(x)$. (*B*) The number of games played over time sliding windows in level 1 in the seven individual regions expressed as a fraction of the total number of games played in these regions in level 1 over the same time window. Here $N = 500$ and the results present an average over 30 simulation runs for *A* and 80 simulation runs for *B*.

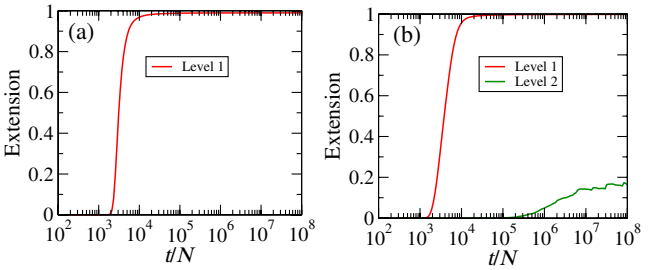


**Fig. S5.** The average extension versus $t/N$ at different levels. (*A*) JND is set to $\bar{d}_{\min}$. (*B*) JND is set to $d_{\min}(x)$ as in Fig. 1 of the main text. Here $N = 500$ and the results present an average over 30 simulation runs.

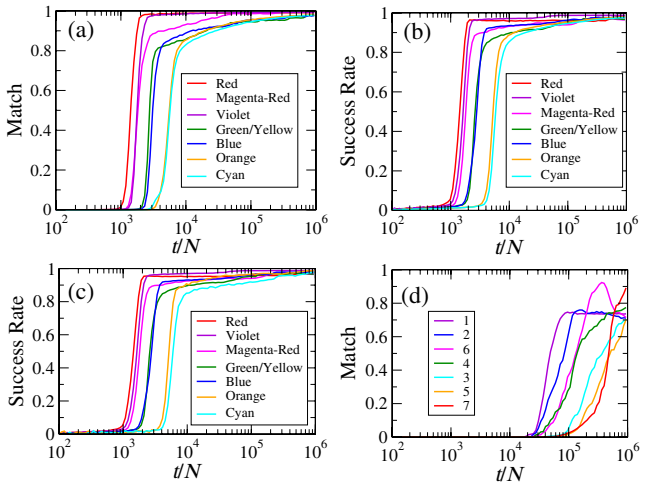


**Fig. S6.** Emergence of regional agreement. (*A*) Match at level 0 in the seven fixed-length regions. (*B*) Success rate in the seven fixed-length regions. (*C*) Success rate in the seven variable-length regions. (*D*) Match at level 1 in the seven variable-length regions. Here $N = 500$ and the results present an average over 30 simulation runs.

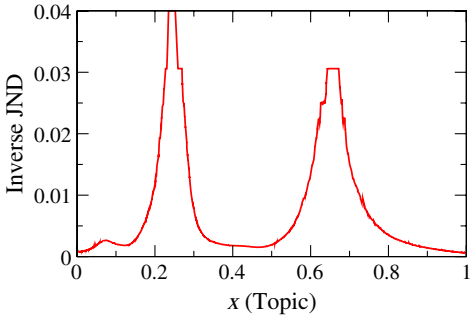


**Fig. S7.** The (properly rescaled) inverse JND function. The function shows the inverse of the actual human JND calculated by first taking $1/d_{\min}(x)$, then normalizing these values by $\max_x[1/d_{\min}(x)]$, and finally multiplying the normalized values with $\max_x[d_{\min}(x)]$, which confines the values in the same range as the actual JND.

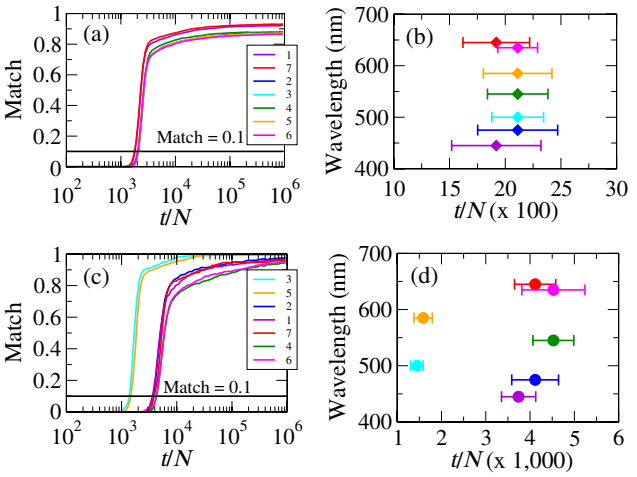


**Fig. S8.** Agreement emergence. Emergence of the agreement in the population in level 0 where the agents are endowed with (*A*) a flat JND and (*C*) an inverse JND. The population size $N = 500$. The time (i.e., $t/N$) for the agents to reach a desired consensus (match = 0.1) versus the value of the wavelength corresponding to the seven regions when they are endowed with (*B*) a flat JND and (*D*) an inverse JND. All the results are averaged over 30 simulation runs.

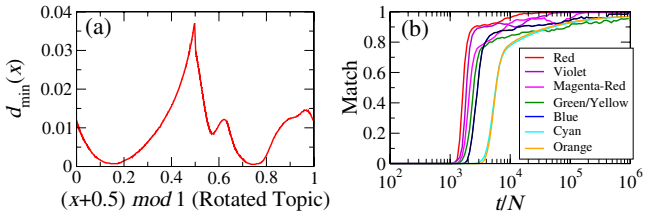


**Fig. S9.** Effect of rotating the stimuli. (*A*) The JND function when the topic values are given a rotation of the form $x \leftarrow (x + 0.5)$ mod1. (*B*) Emergence of the agreement in the population in level 0 where the agents are endowed with the rotated JND function. The population size $N = 500$. All the results are averaged over 45 simulation runs.

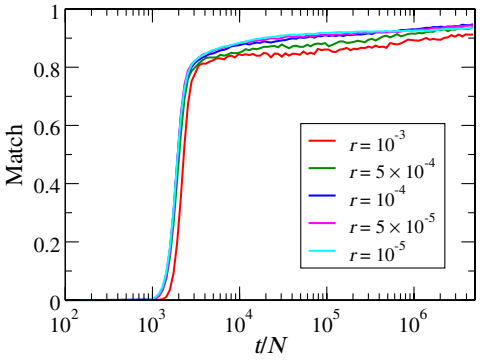


**Fig. S10.** Emergence of the agreement in the population in level 0. We report the overall match for $N = 300$ for different values of the replacement rate $r = 10^{-3}$, $5{\cdot}10^{-4}$, $10^{-4}$, $5{\cdot}10^{-5}$, $10^{-5}$. All the results are averaged over 30 simulation runs.