Supporting Information

Puglisi *et al.* 10.1073/pnas.0802485105

SI Text: Detailed Description of the Rules of the Category Game

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

At each time step $t = 1, 2, \ldots$ 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 that $j \in C \Leftrightarrow j \equiv k$.

There are two possibilities for the topic h:

Either h is already discriminated by a category C;
or h and a nonempty 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 to the hearer a word 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

(*i*) whose inventories contain the transmitted word and
(*ii*) 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 imagine that individuals have a built in conventionalized 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 of 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 of the words but the transmitted one from the inventory of the category discriminating the topic.

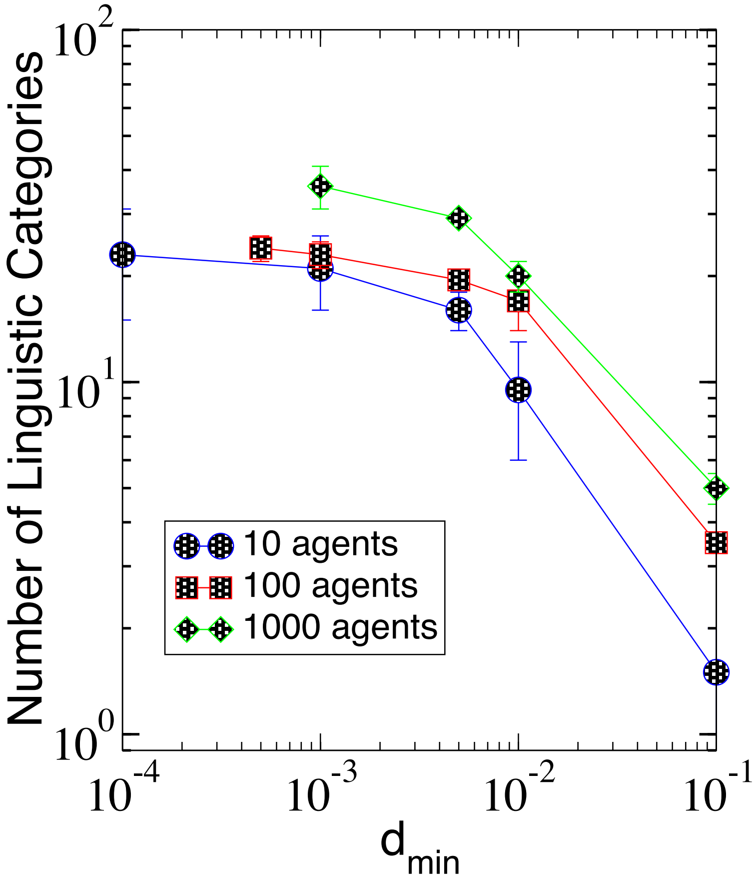

Fig. S1. Effect of the population size. The asymptotic number of linguistic categories as a function of the resolution d_{min} and for different values of the population size N. Data have been taken at time $t = 10^7$ and averaged over different realizations of the simulation; the error bars (sometimes smaller than the symbol size) correspond to the statistical error.

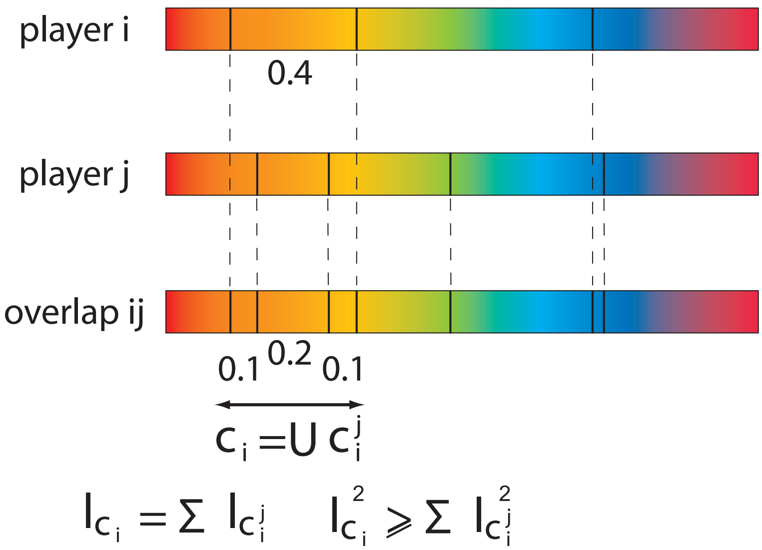

Fig. S2. Example of the calculation of the overlap function between the categories of two agents. Each pair of players *i* and *j* contribute to the total overlap *O* with its pair-overlap o_{ij} defined in Eq. **1**. Here, we give an example of calculation of o_{ij}. In this example, the two players are not perfectly aligned (in particular player *j* has one category more than player *i*). The ''overlap player *ij*'' is constructed, merging all boundaries from both players. Note that if the two players were perfectly aligned, the overlap player would result identical to them. Otherwise, in the overlap player, each category c_i of player *i* appears split into subcategories c_i^j by unaligned boundaries coming from player *j*. And the same happens for categories c_j of player *j*. To make it clear, one example of this splitting is put in evidence in the figure: a category c_i^j of player *i* has length 0.4, and in the overlap player it appears split into three subcategories of length 0.1, 0.2, and 0.1, respectively. The square of the length of c_i, denoted as $l_{c_i}^2$, is always larger than the sum of the squares of its subcategories $l_{c_i^j}^2$. The equals sign holds only when the boundaries are aligned: in that case, it would be $c_i \equiv c_i^j$.