Blue banana – the linguistic phenomenon and not the discontinuous corridor of urbanisation in Western Europe

Elisa Kreiss, Judith Degen, Noah D. Goodman
ekreiss@uos.de, {jdegen,ngoodman}@stanford.edu
Department of Psychology, 450 Serra Mall
Stanford, CA 94305 USA

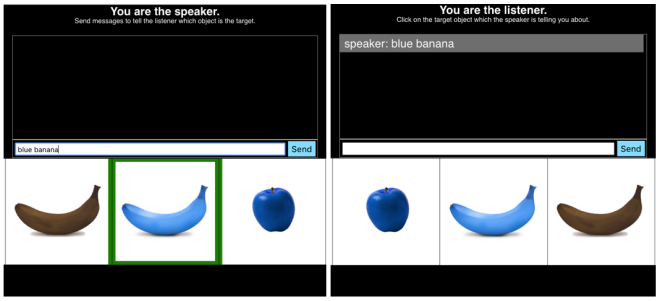

Figure 1: Experimental setup.

Abstract

What we want to find out: Are the typicality values fix in our semantic knowledge, or are they constantly updated by our world knowledge? Therefore, does an exposure of a new frequency of an objects color influence how we produce overinformative expressions?
If typicality values are taken from world knowledge, we have 0 and 1 valued semantics that are then multiplied with knowledge from the real world. If the typicality is in the semantics, we already have truth conditional values between 0 and 1. (Important when thinking about how to construct the model.)
Keywords: keywords

Experiment: color reference game

Methods

Participants and materials We recruited 60 self-reported native speakers of English over Mechanical Turk. The experiment was a multi-player reference game in which one participant was randomly assigned to the role of the speaker, and the other one to the role of the listener. The speaker had to communicate what the target in the context was, and the listener clicked the object they assumed to be the target. The speaker and the listener could communicate freely through a chatbox.

The stimuli were selected from seven food items that each occurred in three different colors, e.g., one of the seven food items was the banana that occurred in the colors yellow, brown, and blue.

Each presented context consisted of three objects, one being the target (the item that had to be referred to), and two being distractors. The contexts were always one out of four possible conditions. The different context types are referred to as "overinformative without a color competitor" (Fig. 2a), "overinformative with a color competitor" (Fig. 2b), "informative without a color competitor" (Fig. 2c), and "informative with a color competitor" (Fig. 2d). An overinformative context describes a situation where mentioning the type of the item, e.g., banana, would be sufficient for an unambiguous identification of the target. An additional mention of color would mean that the speaker uses the color adjective overinformatively, i.e., they are adding "unnecessary" information. However, in this condition the target never has a color competitor, i.e., if the target is brown, there is no distractor of the same color in the context, which means that an only-color utterance would lead to an unambiguous identification, too. This is why there is additionally one overinformative condition with a color competitor (Fig. 2b). In the informative conditions, one has to say the color in addition to the type to make an unambiguous utterance. Again condition, one is without (Fig. 2c) and the other one with a color competitor (Fig. 2d).
The item selection is random but conditioned on the corresponding context condition, i.e., the items need to fulfill the properties dictated by the condition. In the end, each subject sees 42 different contexts. All of the differently colored items are the target for exactly two times but the context in which they occur is drawn randomly from the four possible conditions mentioned above. All in all, we looked at 84 different configurations, i.e., seven food items, each of them in three colors where each could occur in four contexts. The trial order was randomized.

Procedure The participants were randomly formed to pairs and each of them was randomly assigned either the role of the speaker, or the listener. They communicated through a real-time multi-player interface as described in (Hawkins, 2015). The virtual environment of the experiment can be seen in Fig. 1. The speaker and the listener saw the same set of objects but in a randomized order to avoid trivial position-based references such as "the left one". The speaker's task was to tell the listener which of the three displayed objects is the target. The target could be identified by the green border around it. Then the listener could either ask further questions, or immediately click on the object they thought was correct. Afterwards both got a feedback showing whether the right object had been selected by the listener, or not.

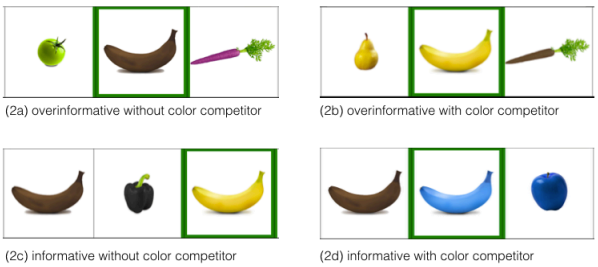

Figure 2: The four context conditions, exemplified by the *banana* domain. The target is outlined in green; the color and type of the distractors differ with each condition (see text).

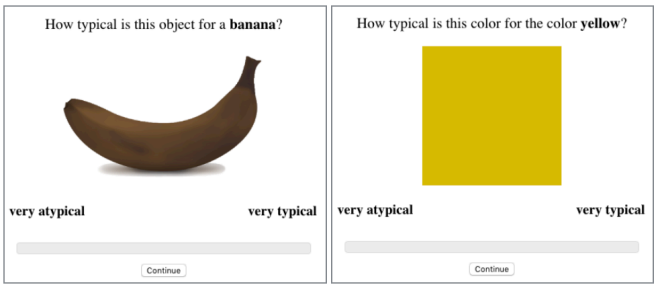

Figure 3: Typicality norming studies for object and color patch norming.

Annotation After collecting the data, the different utterances had to be labeled as belonging to one of the following categories: type-only ("banana"), color-and-type ("yellow banana"), color-only ("yellow"), category-only ("fruit"), and color-and-category ("yellow fruit"). Before sorting, trials have been excluded where the speaker was absent and the listener clicked random objects (the listener self-reported this issue in the comments after completing the study). Additionally, trials were excluded where the listener could not identify the target. The remaining utterances (..% of the original set) had to be cleaned manually from misspellings and abbreviations, e.g., "avoc" for avocado.
Then the resulting utterances were categorized. Still, ..% had to be labeled as "other" with referential expressions, such as "monkeys love..." for banana, or "bugs bunny" for carrot.

Typicality norming For further analysis of the objects, we conducted two norming studies - an object and a color norming study. Both had 75 participants that were recruited over Mechanical Turk.
In the object norming study, the participants were shown a colored food item, e.g., a blue banana, and asked "How typical is this object for an X?" (X being a type, e.g., banana, or a category, e.g., fruit). The participants could rate the fitness on a continuous draggable scale from "very atypical" to "very typical". Every object in the set was paired with every type and category expression from the set, resulting in 189 different trials.
The aim of the color norming study was to identify how typical a certain object's color is for different color terms, i.e., how blue is the given color. The color patches were constructed by determining the average color of the object, stems and leaves excluded. The participants were presented with this color patch, and had to rate its typicality given a certain color term. Again, they had a continuous draggable scale with the end marks "very atypical" and "very typical". Eight color terms were paired with 21 color patches and which resulted in 168 trials.
Due to the amount of trials, each participant only saw a subset of 90. The decision on which trials were shown was random with a slight preference on fitting combinations, e.g., seeing a blue banana and asking for "banana", in contrast to seeing a blue carrot and asking for "apple". The amount of data per combination ranged from 13 to 60. The assumed typicality values are the averaged slider values for each combination ranging from .. (very atypical) to .. (very typical).

Results

Modeling level of reference

Discussion and conclusion

Acknowledgments

This work was supported by ONR grant N00014-13-1-0788 and a James S. McDonnell Foundation Scholar Award to NDG and an SNF Early Postdoc. Mobility Award to JD. RXDH was supported by the Stanford Graduate Fellowship and the National Science Foundation Graduate Research Fellowship under Grant No. DGE-114747.