STANFORD
UNIVERSITY

Qualifying Paper 1

by

Elisa Kreiss

Committee
Chair: Judith Degen
Dan Lassiter
Tobias Gerstenberg

A paper submitted in partial fulfillment
of the requirements for
Candidacy
in
Linguistics

INSERT DATE OF SUBMISSION

Abstract

INSERT SOME ABSTRACT

Contents

[ek: determine how to spell (non)colordiagnosticity] [ek: watch your tenses; define a common textit/quotation style]

1 Experiment: Norming

We want to manipulate the modifier production probabilities a listener can expect from a speaker for an object in isolation. Color modifiers are more likely to be used in isolation (i.e., redundantly) than other adjective types Pechmann and Deutsch (1982). Furthermore, their use is not arbitrary but dependent on the noun they modify Sedivy (2003); Tanaka and Presnell (1999). If one particular color is a defining property for the object, the object is *color-diagnostic* and speakers rarely use the color modifier redundantly to refer to it Tanaka and Presnell (1999). Color-diagnostic objects are for example bananas, which are generally associated with the color yellow. Cups on the other hand are non-color-diagnostic objects, since they are not associated with one particular color and color itself is not a perceptual property that defines it. Even though a sportscar is primarily associated with the color red, color itself is not a defining perceptual property of the object, which is why it is also considered a non-color-diagnostic object Tanaka and Presnell (1999).

Although color-diagnostic objects are rarely modified redundantly when they occur in their typical color, they often are when they occur in an atypical color instead. For example, a yellow banana is mainly referred to as *a banana*, while a blue banana is referred to as *a blue banana* Westerbeek, Koolen, and Maes (2015). To manipulate the modifier production probabilities a listener can expect, we therefore use typical and atypical instances of color-diagnostic objects.

This design posits a number of requirements onto the items used in the study, which is why a variety of objects was carefully normed and then a subset of them was selected for the experiments. In addition to being color-diagnostic (Experiment 1.1), the items had to be shape-diagnostic, such that they would still be recognizable when changing their color. For example, plums, oranges and lemons can barely be told apart when changing their color to something atypical (Experiment 1.4). The items also have to be

known to most participants and be easily recognizable (Experiment 1.2 and 1.4). Since participants will hear the utterance in the eye-tracking experiment, there should only be one potential label for the object to avoid surprisal artifacts at the noun onset [ek: citation?].

Furthermore to our knowledge, previous experiments that manipulated the color typicality of objects, the colors used for typical and atypical instances of objects were not counterbalanced with respect to color, such that for example the color blue primarily occurred as an atypical instance while the color green occurred as a typical one. Since colors vary in their salience and affect eye movement [ek: cite!], this imbalance might be a non-negligible confound to the typicality effect. Each color in our data set occurs twice as a typical instance and twice as an atypical instance.

Finally, the typical and atypical instance of each object was normed to ensure that the color manipulation of the images show the desired difference in typicality ratings (Experiment 1.3).

Motivated through our experimental design [ek: more?], we aimed to find ten color diagnostic objects, evenly distributed over five different colors. To find the most ideal items, we started off with six colors (green, orange, pink, red, white, yellow), each with four possible typical color-diagnostic instances (24 items in total).

1.1 Norming for color-diagnosticity

Firstly, all potential stimuli were normed for their color-diagnosticity. Items are considered color-diagnostic if color is a perceptual property that is closely associated with the object, and if there is agreement on the specific color it is associated with Tanaka and Presnell (1999).

Participants. We recruited 40 participants over Amazon's Mechanical Turk. [ek: add participant recruitment information/payment/duration/...] All participants indicated that their native language was English.

Materials and procedure. The experimental design is adapted from Tanaka and Presnell (1999). Participants were asked to list three perceptual features of an object,

which they entered into three free production text boxes. They could proceed to the next trial if they either entered all three features or indicated by a button press that they did not know the object. In the beginning of the experiment, participants saw an example trial where the term "perceptual feature" was defined and showed an example response for the object *dime*, which was described as *round*, *shiny*, and *small* (as used in Tanaka and Presnell (1999)). Each participant saw 52 trials, four of which were control trials with nonce words. From the remaining trials, 25 asked for presumably color-diagnostic objects (four for each of the six colors and one additional green object), and 23 asked for presumably non-color-diagnostic objects. [ek: Figure: show example design, possibly next to example trial]

Analysis and exclusions. All participants indicated that they were unfamiliar with the four nonce words we included as attention checks. Two participants were excluded because they rated more than eight objects as unknown to them, resulting in a total of 38 participants.

Results. We evaluated the results according to whether a color was mentioned as a first feature, and if a color was mentioned did participants agree on a specific color. The goal is to find five colors, each of which have two highly color-diagnostic objects.

First of all, color was more likely to be mentioned for the objects that were intended to be color-diagnostic and less likely for the non-color-diagnostic ones [ek: Figure]. An exception to that is bell pepper which is expected to be non-color-diagnostic but for which participants still mention color as an important perceptual feature. However bell pepper is not only associated with one specific color. Although it was mainly described as *green*, several participants also mentioned *red*. For all items that were intended to be color-diagnostic it holds that, if a color was mentioned, participants generally agreed on the color.

Figure [ek: ref] shows the proportion of color mention as the first perceptual feature, facetted by the color they were associated with. The dashed horizontal line marks the case when color was mentioned half of the time as the first perceptual feature. The items in the colors pink and white have the lowest proportion of color mention overall,

suggesting that most likely one of these colors will not be used for the final data set. Items will low proportions of color mention in each color are less likely to be chosen as stimuli.

1.2 Norming for nameability

After we normed for color-diagnosticity, we chose image depictions of the items and normed them for their nameability.

Participants. We recruited 20 participants over Amazon's Mechanical Turk. [ek: add participant recruitment information/payment/duration/...] All participants indicated that their native language was English.

Materials and procedure. Each participant saw 50 trials in which they were asked "What is this?" with a depiction of an object and a free production text field. We used the same objects as in the color-diagnosticity norming study (i.e., 25 color-diagnostic objects and 23 non-color-diagnostic ones), but added an additional depiction for the lettuce and the sportscar.

Exclusions. Two participants were excluded because they indicated that they were unsure whether they did the experiment correctly, resulting in a total of 18 participants.

Results. We evaluated the results according to how many labels were used. If more than one label was used, we favored cohort competitors over entirely separate terms (e.g., *bike* and *bicycle* are more acceptable deviations than *traffic cone* and *cone*). Overall, participants agreed on the label to give to the different objects. However there were also some problematic items. Both, the pickle and zucchini, were called cucumbers and for the zucchini this even happened to the same degree as the actual label. Given that they also have a low shape-diagnosticity, presenting them in atypical colors will most likely even worsen these issues, which is why they are candidates to exclude. Other items that received a variety of labels were the traffic cone (e.g., *traffic cone, cone, caution cone, hazard cone*) and the rubber duck (e.g., *rubber duck, duck, duck toy*). These cases are problematic because the different labels are not cohort competitors of each other, which will be relevant for the eyetracking experiments.

Finally, we investigated which depiction represents the generic lettuce the best: romaine or iceberg. We found that when participants saw the iceberg lettuce before the romaine lettuce, they simply called it *lettuce*. However, if they saw the romaine lettuce first, they called it *romaine lettuce* 20% of the time. This suggests that the iceberg lettuce is the more prototypical lettuce (in the MTurk community).

1.3 Norming for typicality

Participants. We recruited 30 participants over Amazon's Mechanical Turk. [ek: add participant recruitment information/payment/duration/...] All participants indicated that their native language was English.

Materials and procedure.

Results.

1.4 Norming for free production

1.5 Norming for multiple choice

1.6 Conclusion

2 Rational Speech-Act Model

3 References

Pechmann, T., & Deutsch, W. (1982). The development of verbal and nonverbal devices for reference. *Journal of experimental child psychology*, *34*(2), 330–341.

Sedivy, J. C. (2003). Pragmatic versus form-based accounts of referential contrast: Evidence for effects of informativity expectations. *Journal of Psycholinguistic Research*, *32*(1), 3–23.

Tanaka, J. W., & Presnell, L. M. (1999). Color diagnosticity in object recognition. *Perception & Psychophysics*, *61*(6), 1140–1153.

Westerbeek, H., Koolen, R., & Maes, A. (2015). Stored object knowledge and the production of referring expressions: the case of color typicality. *Frontiers in Psychology*, *6*.