



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

1	Introduction	1
1.1	SOME SUBSECTION	1
2	Experiment: Norming	1
2.1	Norming for color-diagnosticsity	1
2.2	Norming for nameability	1
2.3	Norming for typicality	2
2.4	Norming for free production	3
2.5	Final stimuli selection	3
3	Experiment: Production	3
3.1	Method	3
3.2	Results	4
4	Experiment: Comprehension	4
4.1	Method	4
4.2	Results	4
5	Discussion	4
5.1	Conclusion	5
6	Appendix	6
7	Acknowledgments	7
8	References	8

[ek: determine how to spell (non)colordiagnosticity]

1 Introduction

1.1 SOME SUBSECTION

2 Experiment: Norming

Started off with 6 colors (green, orange, pink, red, white, yellow), each with 4 possible colordiagnostic instances; goal was to find 2 cd objects for 5 of the colors, because we always need at least two typical instances of a color to create the tt(p/n) conditions;

2.1 Norming for color-diagnosticity

Inspired by Tanaka and Presnell

Goal: determine color diagnosticity (1) Is color a property that is closely associated with the object? (2) If a color is mentioned, do participants agree on the color?

Task: "List 3 perceptual features of a **NOUN**"

free production and checkbox with possibility to say "I don't know this object."

52 trials; 4 control trials with nonce words; 25 presumably color diagnostic objects (4 for each of the 6 colors + 1 more green thing) and 23 presumably non-color diagnostic objects

40 participants; exclusion criteria: everyone correctly identified the 4 nonce words as unknown objects; and if they rated more than 8 objects as "object unknown" they were excluded (2 participants); resulting number of participants: 38

we evaluated the results according to whether (1) a color was mentioned at all in the features (2) a color was mentioned as a first feature (3) if a color was mentioned, was it the same or did they differ

2.2 Norming for nameability

Goal: Are the image depictions we chose nameable, the way we intended?

Task: "What is this?"

free production

50 trials; 26 depictions of presumably color diagnostic objects (same as in color diagnosticity norming + 1 more lettuce depiction) and 24 presumably non-color diagnostic ones (same as in color diagnosticity norming + 1 more (sports)car); 20 participants; exclusion: 2 participants because they indicated that they were confused or didn't do the HIT correctly; resulting number of participants: 18

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 than traffic cone and cone); Wrt to lettuce we had romaine and iceberg lettuce depictions. The simple noun lettuce was more frequently used for the iceberg than the romaine lettuce which is why we favored the iceberg lettuce.; other things: zucchini was half of the time misclassified as cucumber, similarly for pickle, traffic cone had a lot of different labels, such as simply cone, caution cone, hazard cone, safety cone

2.3 Norming for typicality

Goal: Does the color manipulation of the images show the desired difference in typicality ratings?

Task: "How typical is this object for a **NOUN**?"

slider rating, underlyingly coded as ranging from 0 to 100

45 trials; 11 color diagnostic objects, each in their typical color and 1-2 atypical colors (i.e., 25 stimuli); 20 non-color diagnostic stimuli

30 participants; exclusions: none; everyone thought they did the HIT correctly

Results: generally clear distinction between typical and atypical instance; From the three items that were normed in two atypical colors (carrot, corn, pumpkin), we see the biggest difference between the red and white pumpkin. Therefore, we should choose the white pumpkin and (following from that) the green carrot and red corn. There does not seem to be a big difference between the yellow egg and snowman, but the white egg is rated even more typical and its size fits better to the other stimuli. Therefore, we should choose the egg over the snowman (given that both are also nameable). Even

though the orange banana is predominantly rated below 50, it is still not as atypical as other objects.; The non-color diagnostic objects are all rated as very typical instances.

2.4 Norming for free production

Goal: Are the image depictions we chose nameable, the way we intended?

Task: "What is this?"

free production

31 trials (22 cd – each participant saw one instance of each object at random, i.e., either typical or atypical; 20 non-cd)

Results: swan is often called a goose; two people identified the white carrot as parsnip

2.5 Final stimuli selection

In the end, we have 10 objects, each occurring in a typical and atypical color. Items can occur in the colors yellow, red, green, orange and white. Each color occurs twice as typical and twice as atypical. This counterbalance aims to reduce artifacts of salience such as red is generally more salient as a warn signal [ek: ref?] and blue is highly atypical for most objects. A full list of stimuli can be found in table [ek: add table and reference].

3 Experiment: Production

3.1 Method

Participants.

Procedure.

Materials.

Data Preprocessing and exclusion.

3.2 Results

4 Experiment: Comprehension

4.1 Method

Participants. We recruited 80 participants over Amazon’s Mechanical Turk, 40 for each color competitor typicality manipulation. The study took on average 7 minutes and each of them were paid [ek: ...] for their participation. We restricted participation to workers with IP addresses in the US and a approval rate of previous work above 97%.

Procedure. This experiment is a one-player adaption of the production study explained in [ek: ref to section]. In the four practice trials, the participant was given the role of the speaker and in the main experiment the role of the listener. In the speaker role, they saw a grid of four non-color diagnostic objects, one of which was marked as the target by a green border surrounding it. They were then asked to refer to the object such that a second player could identify it. The practice trials were introduced to familiarize the participants with the task.

In the main part of the experiment, participants were put into the listener role of the reference game. That means, they needed to identify which object was the target given a referring expression placed above the grid. Crucially, they do not get the complete referring expression at once. The utterance is gradually revealed to the parti...[ek: CONTINUE HERE] (This method is highly inspired by the incremental decision task introduced by [ek: cite Qing].)

4 practice trials from the speaker perspective with non-color diagnostic objects. Then 55 trials (20 critical, 35 fillers).

Materials.

Data Preprocessing and exclusion.

4.2 Results

5 Discussion

Summary.

5.1 Conclusion

6 Appendix

The appendix could not be included, due to file size restrictions. If you are interested in the full version of the thesis, I am happy to provide it.

7 Acknowledgments

First and foremost, I would like to thank Judith Degen for her scientific guidance in all respects and also for her personal support throughout the whole project.

I would also like to thank the Computation and Cognition Lab at Stanford University for hosting me, which enabled this project in the first place. I am especially thankful to Robert X. D. Hawkins and Noah D. Goodman who were always available for scientific and personal advice and discussions.

Finally, I'm grateful for the continuous support from Eleni Gregoromichelaki who provided place for project-related exchanges with other students and whose belief in me was a constant motivation.

8 References