Whorfian colours: a series of pre-registered experiments on the effect of verbal labels on perception

[G. Montero-Melis, M. Ostarek, J. van Paridon]

# Introduction: aim and core questions

Can verbal labels shape our perception of the world? This question lies at the heart of the much-debated linguistic relativity hypothesis that language shapes thought (Boroditsky, 2012; Gumperz & Levinson, 1996; McWhorter, 2014; Pinker, 1994; Whorf, 1956; Wolff & Holmes, 2011). Surprisingly, evidence *directly* addressing this question remains limited. Here we implement a training study in which participants are trained to distinguish novel colour categories by associating them to novel linguistic labels during a multi-session learning phase. Our design controls for exposure to different colours, making category assignment the only relevant difference between the colours. A causal effect of language can be *directly* tested by comparing how well participants discriminate colours that belong to the same or to a different category before and after training. Our design controls for inter-stimulus differences by having the same pairs of colours belong to the same category for one group and to different categories for another group.

In a series of incremental pre-registered experiments, we aim to answer three/four [?] core questions about the effect of linguistic labelling on perception with colour as the test domain.

1. *Can* verbal categories boost the ability to perceptually discriminate differences between colours? We are interested in whether training results in improved discrimination in two tasks:
   1. *Memory task*: A task in which the colours to be discriminated are shown sequentially to the observer and which therefore involves keeping the colours in short-term memory.
   2. *Pure discrimination task*:A task in which the colours to be discriminated are shown simultaneously to the observer and which therefore does not involve memory.
2. Does the strength (or even the presence) of the effect depend on using *verbal labels* as opposed to other equally informative cues that are non-language-like?
3. Is the effect mediated by online access to the labels? If so, verbal interference should disrupt the effect but equally demanding non-verbal interference should not.
4. [Is the effect due to early perceptual changes? If so, then the training should not transfer to other regions of the colour space (for the logic of this inference, see Ahissar & Hochstein, 2004; Notman et al., 2005).]

# Shortcomings of cross-linguistic designs

Traditionally, whether language affects perception has been addressed with cross-linguistic designs – and the domain of colour has served as the primary test bed. However, such experimental designs fail to provide a strong direct test of the effect of linguistic categories on perceptual processes. First, differences between languages may covary with other cultural or environmental factors, making it difficult to isolate effects of language per se. For the same reason, correlations between linguistic and perceptual differences cannot license inferences about causality, as other cultural differences could in principle be the cause of both linguistic and perceptual differences.

Second, although cross-linguistic differences in the number of basic colour terms are well documented (Berlin & Kay, 1969; Kay et al., 2009), these differences do not appear to be unconstrained, but have been argued to follow universal principles possibly determined by physiological constraints of the human visual system (Berlin & Kay, 1969; but see Witzel, 2019). If so, cross-linguistic differences become poor candidates to test the effect of language on perception, as they would be biased towards emphasizing differences that are perceptually salient independent of language.

Third, naturally occurring colour categories are fuzzy: between-speaker agreement is high for some regions of the colour space (i.e., a prototypical blue or green), but low around category boundaries (some hues will be considered blue or green by different speakers) (Lindsey & Brown, 2014). Yet it is precisely along category boundaries that perceptual effects of categories are expected to emerge (Goldstone & Hendrickson, 2010). Since colour category boundaries are not systematically disambiguated through language, this weakens the potential effect one might find in cross-linguistic designs.

# Basic experimental design

Our basic design involves training participants on the following novel categorical distinctions:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Group 1 | L1 | | L2 | |  |
| C1 | C2 | C3 | C4 |  |
| Group 2 |  | C2 | C3 | C4 | C5 |
|  | L1 | | L2 | |

Figure 1. Basic experimental design. C1, C2, etc. stand for equidistant colours differing only along the hue dimension in a given region of colour space (the colours are only illustrative and not the real stimuli used in the study). Group 1 and Group 2 are two distinct groups of participants. Each group is trained to associate two adjacent colours with one label (L1) and two further adjacent colours with another label (L2). Across groups, the same pair of colours (e.g., C2-C3) either belong to different categories (Group 1) or to the same category (Group 2). In the test phase, colour discrimination is compared for colours belonging to the same or different categories.

Our experimental design rules out that increased perceptual discrimination is due to something else than the labels. This is achieved by counterbalancing the stimuli and conditions, such that the same two adjacent stimuli (e.g., C2-C3) belong to different categories for half of the participants (Group 1), but to the same category for the other half (Group 2).

# Overview of experiments and predictions

Given the sequential nature of the questions asked, our experiments will be run sequentially. Each experiment is motivated only if the previous one confirms our predictions. This means that an initial failure to find any effect of linguistic labels on perception would remove the necessity to run subsequent studies. Here we lay out our general predictions and procedures. We will pre-register the details of each experiment in turn before running it.

## Experiment 1:

## Experiment 2:

## Experiment 3:

# References

Ahissar, M., & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends in Cognitive Sciences*, *8*(10), 457–464. https://doi.org/10.1016/j.tics.2004.08.011

Berlin, B., & Kay, P. (1969). *Basic color terms: Their universality and evolution*. University of California Press.

Boroditsky, L. (2012). How the languages we speak shape the ways we think: The FAQs. In M. Spivey, K. McRae, & M. Joanisse (Eds.), *The Cambridge Handbook of Psycholinguistics* (pp. 615–632). Cambridge University Press.

Goldstone, R. L., & Hendrickson, A. T. (2010). Categorical perception. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(1), 69–78. https://doi.org/10.1002/wcs.26

Gumperz, J. J., & Levinson, S. C. (Eds.). (1996). *Rethinking linguistic relativity* (Vol. 17). Cambridge Univ. Press.

Kay, P., Berlin, B., Maffi, L., Merrifield, W. R., & Cook, R. (2009). *The world color survey*. CSLI Publications.

Lindsey, D. T., & Brown, A. M. (2014). The color lexicon of American English. *Journal of Vision*, *14*(2), 17–17. https://doi.org/10.1167/14.2.17

McWhorter, J. H. (2014). *The language hoax: Why the world looks the same in any language*. Oxford University Press.

Notman, L. A., Sowden, P. T., & Özgen, E. (2005). The nature of learned categorical perception effects: A psychophysical approach. *Cognition*, *95*(2), B1–B14. https://doi.org/10.1016/j.cognition.2004.07.002

Pinker, S. (1994). *The language instinct: The new science of language and mind*. Penguin.

Whorf, B. L. (1956). Science and linguistics [1940]. In J. B. Carroll (Ed.), *Language, thought, and reality. Selected writings of Benjamin Lee Whorf* (pp. 207–219). MIT Press.

Witzel, C. (2019). Misconceptions About Colour Categories. *Review of Philosophy and Psychology*, *10*(3), 499–540. https://doi.org/10.1007/s13164-018-0404-5

Wolff, P., & Holmes, K. J. (2011). Linguistic relativity. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(3), 253–265. https://doi.org/10.1002/wcs.104