

The Sapir-Whorf Hypothesis and Probabilistic Inference: Evidence from the Domain of Color

Emily Cibelli¹ (emily.cibelli@northwestern.edu)

Yang Xu^{2,3} (yang_xu_ch@berkeley.edu)

Joseph L. Austerweil⁴ (joseph_austerweil@brown.edu)

Thomas L. Griffiths^{3,5} (tom_griffiths@berkeley.edu)

Terry Regier^{2,3} (terry.regier@berkeley.edu)

¹ Department of Linguistics, Northwestern University, Evanston, IL 60208 USA

² Department of Linguistics, University of California, Berkeley, CA 94720 USA

³ Cognitive Science Program, University of California, Berkeley, CA 94720 USA

⁴ Department of Cognitive, Linguistic, and Psychological Sciences, Brown University,
Providence, RI 02912 USA

⁵ Department of Psychology, University of California, Berkeley, CA 94720 USA

Keywords: Sapir-Whorf hypothesis; linguistic relativity; color memory; category effects; probabilistic inference.

The Sapir-Whorf hypothesis holds that our thoughts are shaped by our native language, and that speakers of different languages therefore think differently. This hypothesis is controversial in part because it appears to deny the possibility of a universal groundwork for human cognition, and in part because some findings held to support it have not reliably replicated. We argue that considering this hypothesis through the lens of probabilistic inference has the potential to resolve both issues, at least with respect to certain prominent findings concerning color cognition.

We cast color memory as inference under uncertainty, and explore this idea using a version of the category adjustment model of Huttenlocher et al. (1991) (see also Bae et al., 2015; Persaud & Hemmer, 2014). The model holds that color memory involves the probabilistic combination of evidence from two sources: a fine-grained representation of the specific color seen, and the language-specific category in which it fell (e.g. *green*). Both sources of evidence are represented in a universal perceptual color space, yet their combination yields language-specific bias patterns in memory. The model predicts that such category effects will be strongest when perceptual information is uncertain. It thus has the potential to explain the mixed pattern of replications of Whorfian effects in the literature (e.g. Brown et al., 2011; Wright et al., 2015): non-replications could result from high perceptual certainty.

We first describe the results of an experiment showing language-consistent biases in color memory in English speakers, and stronger bias when perceptual information is uncertain, consistent with this model, and replicating recent findings (Bae et al., 2015). We then show that the same model also accounts for existing data on cross-language differences in color memory in speakers of English, Berinmo, and Himba (Roberson et al., 2000; 2005). Finally, we show that this model accounts for existing cross-language data on within-category discrimination (Hanley &

Roberson, 2011). We suggest that these ideas may help to clarify the debate over the Sapir-Whorf hypothesis.

Acknowledgments

We thank Paul Kay and Charles Kemp for their comments. This research was supported by NSF under grants DGE-1106400 (EC) and SBE-1041707 (YX, TR).

References

- Bae, G.Y., Olkkonen, M., Allred, S.R., & Flombaum, J.I. (2015). Why some colors appear more memorable than others: A model combining categories and particulars in color working memory. *Journal of Experimental Psychology: General*, 144, 744-763.
- Brown, A.M., Lindsey, D.T., Guckes, K.M. (2011). Color names, color categories, and color-cued visual search: Sometimes, color perception is not categorical. *Journal of Vision*, 11, 2.
- Hanley, J.R. & Roberson, D. (2011). Categorical perception effects reflect differences in typicality on within-category trials. *Psychonomic Bulletin and Review*, 18, 355-363.
- Huttenlocher, J., Hedges, L.V., & Duncan, S. (1991). Categories and particulars: Prototype effects in estimating spatial location. *Psychological Review*, 98, 352-376.
- Persaud, K. & Hemmer, P. (2014). The influence of knowledge and expectations for color on episodic memory. In *Proceedings of the 36th Annual Meeting of the Cognitive Science Society*, 1162-1167.
- Roberson, D., Davies, I., & Davidoff, J. (2000). Color categories are not universal: Replications and new evidence from a stone-age culture. *Journal of Experimental Psychology: General*, 129, 369-398.
- Roberson, D., Davidoff, J., Davies, I.R.L., & Shapiro, L.R. (2005). Color categories: Evidence for the cultural relativity hypothesis. *Cognitive Psychology*, 50, 378-411.
- Wright, O., Davies, I.R.L., & Franklin, A. (2015). Whorfian effects on colour memory are not reliable. *The Quarterly Journal of Experimental Psychology*, 68, 745-758.