LEARNING TO COMMUNICATE
ABOUT CONCEPTUAL HIERARCHIES

Robert X. D. Hawkins[1*], Kenny Smith[2], Michael Franke[3], and Noah D. Goodman[1,4]

[*]Corresponding Author: rxdh@stanford.edu
[1]Department of Psychology, Stanford University, USA
[2]Centre for Language Evolution, University of Edinburgh, UK
[3]Department of Linguistics, University of Tübingen, Germany
[4]Department of Computer Science, Stanford University, USA

Natural languages provide speakers with remarkable flexibility in the labels they may use to refer to things (Brown, 1958). In addition to the combinatorial explosion of modifiers afforded by compositionality (Partee, 1995), we have a number of lexicalized nominal terms at our disposal. *Dalmatian*, *dog*, and *animal* can all truthfully be used to talk about the same Dalmatian at different levels of specificity, with one level of the conceptual hierarchy – the basic-level – generally privileged over the others (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). How these overlapping meanings are learned, and why speakers choose different levels of specificity in different contexts, is increasingly accounted for by probabilistic models of pragmatic language use (e.g. Xu & Tenenbaum, 2007; Graf, Degen, Hawkins, & Goodman, 2016) but there remains a more fundamental evolutionary question: how do multiple levels of reference come to coexist in the lexicon to begin with?

Our hypothesis, motivated both by classic work on concept representations and contemporary work on the selective pressures induced by communication, is that lexicalization of conceptual hierarchies is a function of (1) the structure and statistics of entities in the environment, and (2) the particular contexts in which communication occurs. In particular, we expect hierarchical lexica to form when features can be encoded as predictable clusters and communicative goals require distinctions to be drawn at multiple levels. To test this hypothesis, we designed a repeated reference game in which pairs of participants interactively created an artificial language to communicate with each other about objects in context (e.g. Winters, Kirby, & Smith, 2014; Galantucci & Garrod, 2011).

In this game, participants were paired over the web and placed in a shared environment containing a grid of four objects (Fig. 1A) and a 'chatbox' to send messages from a pre-specified vocabulary of sixteen words (Fig. 1B). On each of ninety trials, one player — the 'speaker' — was privately shown a highlighted target object and allowed to send a single word to help their partner select this object
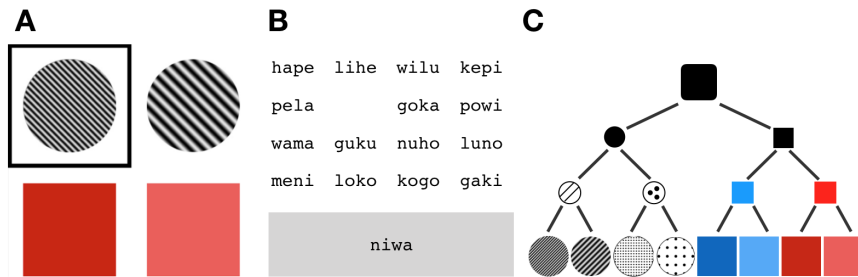
Figure 1. (A) Example array of elements the matcher must choose from. The target is highlighted for the speaker with a black square. In this *subordinate-required* trial there is a distractor at the same intermediate level (striped circle) as the target, so using any abstract label would be insufficient. (B) Drag-and-drop chat box interface. (C) Hierarchical organization of stimuli.

from the array of distractors. The set of objects was designed to cluster in a fixed three-level hierarchy (Fig. 1C). Distractors could differ from the target at any of the levels, creating three kinds of contexts defined by the finest distinction that had to be drawn. In addition to behavioral trajectories observed over the course of the game, we conducted a post-test to explicitly probe players' lexica. For each word, we asked players to select all objects to which that word applies, allowing us to distinguish between subordinate terms that apply to only one element and abstract terms that apply to multiple elements at an intermediate level of the hierarchy: striped circles, for example.

Critically, we manipulated the statistics of the context in a between-subjects design to test the contribution of communicative relevance to lexicalization. In one condition, all three kinds of context were equally likely, thus providing high diversity in the relevant distinctions that must be drawn. We also ran three control conditions in which a single kind of context dominated, e.g. in the 'subordinate-required' condition, the majority of trials contained distractors that were close neighbors to the target (e.g. Fig. 1A), theoretically requiring speakers to lexicalize a label for each target.

We counted the relative number of subordinate-level terms and abstract terms in the post-test and found that the likelihood of lexicalizing abstractions differed significantly across conditions; in particular, the uniform condition was more likely to give rise to lexica in which multiple levels of reference coexist. This suggests that pragmatic pressures for informativity in a diversity of communicative contexts is instrumental for the lexicalization of hierarchical reference systems. Our separate minds may organize the world into meaningful conceptual hierarchies but our shared language only evolves to reflect this structure when it is communicatively relevant.

References

Brown, R. (1958). How shall a thing be called? *Psychological review*, *65*(1), 14.

Galantucci, B., & Garrod, S. (2011). Experimental semiotics: a review. *Frontiers in human neuroscience*, *5*, 11.

Graf, C., Degen, J., Hawkins, R. X. D., & Goodman, N. D. (2016). Animal, dog, or dalmatian? level of abstraction in nominal referring expressions. In A. Papafragou, D. Grodner, D. Mirman, & J. Trueswell (Eds.), *Proceedings of the 38th annual conference of the Cognitive Science Society*.

Partee, B. (1995). Lexical semantics and compositionality. In L. R. Gleitman & M. Liberman (Eds.), *An invitation to cognitive science, part i: Language* (pp. 311–360). Cambridge, MA: MIT Press.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive psychology*, *8*(3), 382–439.

Winters, J., Kirby, S., & Smith, K. (2014). Languages adapt to their contextual niche. *Language and Cognition*, 1–35.

Xu, F., & Tenenbaum, J. B. (2007). Word learning as bayesian inference. *Psychological review*, *114*(2), 245.