An inferential theory of convention formation

Authors
Universities

Abstract

Keywords: keywords

Experiment 2: generalization to new partners in a social network

How do *ad hoc* conventions formed through interaction with a single partner become *global* conventions shared throughout a community? In this section, we provide an explicit computational account of the cognitive mechanisms driving this shift. The key predictions distinguishing our model concern the pattern of generalization across partners. First, we show that our model accounts for the *partner-specificity* of ad hoc conventions as a consequence of hierarchical structure. Under our model, speakers revert back to a longer description with a novel partner because evidence from a single listener is relatively uninformative about the community-level prior.

This hierarchical structure, however, leads to a further prediction: after interacting with enough partners in a tight-knit community, speakers should become increasingly confident that labels are not simply idiosyncratic features of a particular partner's lexicon but are shared across the entire community. In other words, the partner-specific expectations agents form within an interaction to solve a novel communication problem should gradually generalize to community-wide expectations as they gain additional evidence of the latent population-level distribution from which different partners are sampled. These expectations should manifest in an increasing willingness to use short labels with novel partners, leading to the emergent consequence of lending additional evidence for that structure. We test this novel prediction in a networked communication game and compare our model to two non-hierarchical variants: a 'complete-pooling' model that collapses across partners and a 'no-pooling' model that treats partners as entirely separate.

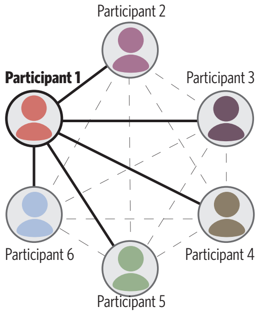

Figure 1. Network

Participants

We recruited 600 participants from Amazon Mechanical Turk to play an interactive, natural-language reference game using the Dallinger platform[1]. Participants were randomly assigned to one of 100 six-member networks and embedded as a fully-connected graph, such that everyone was neighbors with everyone else in their assigned community (Fig. 1). We prescreened participants who report a native language different from English.

Stimulus & design

Participants were paired with each of their five neighbors in a round-robin schedule. In each interaction, they played a game where they repeatedly referred to six

This report is based in part on work presented at the 38th Conference of the Cognitive Science Society. Correspondence should be addressed to Robert D. Hawkins, e-mail: rdhawkins@princeton.edu

[1] http://docs.dallinger.io/

abstract tangram shapes taken from Clark and Wilkes-Gibbs (1986, Fig. 2). These stimuli have been used extensively in the literature on coordination and common ground (e.g. Duff et al., 2006; Hawkins et al., 2017). They were designed such that participants will not already have strong pre-existing lexical conventions for how to refer to them (unlike photographs of common objects), but are structured enough to support many possible descriptions (unlike images of white noise).

On each trial of a reference game, one of these six shapes was highlighted as the *target object* for the "speaker" who was instructed to use a chatbox to communicate the identity of this object to their partner, the "listener". The listener may reply through the chatbox but must ultimately make a selection from the array. The trial sequence for a given partner was constructed so that each of four targets appear six times each, spread evenly across the session, for a total of 24 trials.

After completing 24 trials with one partner, they were introduced to their next partner and asked to play the repeated reference game again with the same four objects. Each participant in a network was assigned a distinct avatar so that participants were clear they were speaking to distinct partners. This process repeated until each participant had partnered with all five neighbors. Players were given full feedback on each about their partner's choice and received bonus payment for each correct response. Because some pairs within the network took longer than others to complete the trial sequence, we sent participants to a temporary waiting room if their next partner was not yet ready.

A remaining design decision is how to handle disconnections in the middle of the game. do we immediately terminate the entire group as soon as one person drops out? Do we just continue the game without them, making whoever was supposed to be their partner wait a long time? do we recruit to fill their place? my first thought is to just terminate the group to minimize turker frustration (i.e. more will probably dropout if they have to wait because of the first dropout) and to avoid having to account for newly recruited turkers in the analysis

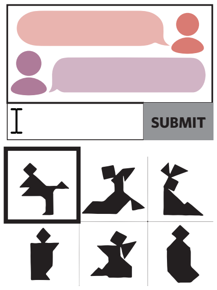

Figure 2. Task

Confirmatory behavioral predictions

The most diagnostic dependent variable to distinguish the hierarchical model from alternatives concerns the mean number of words used per description. This is a common operationalization of conventionalization in terms of coding efficiency. In particular, we are interested in how this measure changes within partners and across partner boundaries. The first panel of Fig. 3 shows the mean number of words used across six repetitions with one partner in a pilot experiment with $N = 100$ isolated pairs. All of the model variants we consider are consistent with this pattern: messages reduce in length across repetitions with a partner as they coordinate on shorthand.

However, our account diverges in its predictions at the partner boundary, as visualized in the remaining panels of Fig. 3. Pilot data was used to calibrate expected effect sizes. While 'complete-pooling' predicts that participants will completely transfer conventions from the previous partner (inconsistent with existing evidence from psycholinguistics, e.g. Wilkes-Gibbs & Clark, 1992), the other accounts predict that speakers will revert nearly to their initial description length. Furthermore, while *no-pooling* predicts the same complete reversion with every subsequent partner, our hierarchical *partial pooling* account predicts that the magnitude of the reversion will decrease over successive interactions: after several partners, it predicts transfer as strongly as complete pooling.

We will test these predictions using a mixed-effects regression of partner number on "reversion size" (the difference in number of words between the final descriptions on one repetition and the initial ones on the next), with maximum random-effect structure including item-effects at the object and speaker level. Only the partial pooling model predicts a significant decrease in reversion over successive partners. Preliminary support for this signature was reported by Fay et al. (2010) in a Pictionary task where participants used sketches to

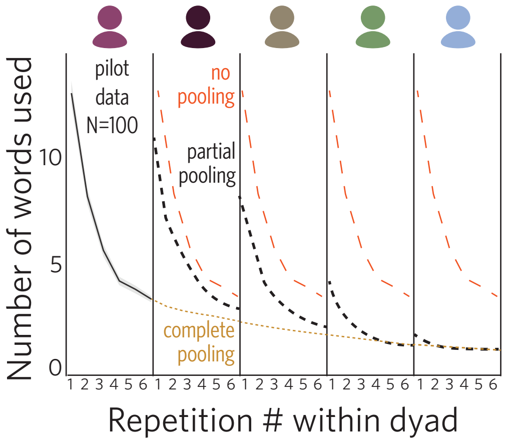

Figure 3. Predictions

communicate verbal concepts instead of using words to refer to visual targets, and the measure of interest was the complexity of the drawings (see also Garrod & Doherty, 1994).

Simulations

Acknowledgments

References

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, *22*(1), 1–39.

Duff, M. C., Hengst, J., Tranel, D., & Cohen, N. J. (2006). Development of shared information in communication despite hippocampal amnesia. *Nature neuroscience*, *9*(1), 140–146.

Fay, N., Garrod, S., Roberts, L., & Swoboda, N. (2010). The interactive evolution of human communication systems. *Cognitive Science*, *34*(3).

Garrod, S., & Doherty, G. (1994). Conversation, coordination and convention: An empirical investigation of how groups establish linguistic conventions. *Cognition*, *53*(3).

Hawkins, R. X. D., Frank, M. C., & Goodman, N. D. (2017). Convention-formation in iterated reference games. In *Proceedings of the 39th annual meeting of the cognitive science society.*

Wilkes-Gibbs, D., & Clark, H. H. (1992). Coordinating beliefs in conversation. *Journal of memory and language*, *31*(2), 183–194.