Modeling contextual flexibility in visual communication

Judith E. Fan[a], Robert X.D. Hawkins[a], Mike Wu[b], and Noah D. Goodman[a,b]

[a]Department of Psychology, Stanford University
[b]Department of Computer Science, Stanford University

December 12, 2017

Visual modes of communication are ubiquitous in modern life — from maps to data plots to political cartoons. Here we investigate drawing, the most basic form of visual communication. Communicative drawing poses a core challenge for theories of how vision and social cognition interact, requiring a detailed understanding of how sensory information and social context jointly determine what information is relevant to communicate. Participants (N=192) were paired in an online environment to play a drawing-based reference game. On each trial, both participants were shown the same four objects, but in different locations. The *sketcher's* goal was to draw one of these objects — the target — so that the *viewer* could pick it out from a set of distractor objects. There were two types of trials: *close*, where objects belonged to the same category, and *far*, where objects belonged to different categories. We found that people exploited information in common ground with their partner to efficiently communicate about the target: on far trials, sketchers achieved 99.7% recognition accuracy while applying fewer strokes, using less ink, and spending less time (ps<0.001) on their drawings than on close trials. We hypothesized that humans excel at this task by recruiting two core competencies: (1) **visual abstraction**, the capacity to perceive the correspondence between an object and a drawing of it; and (2) **social reasoning**, the ability to infer what information would help a viewer distinguish the target from distractors. We instantiated these competencies in a computational model that combines a multimodal convnet visual encoder with a Bayesian model of recursive social reasoning, and found that it fit the data well and outperformed lesioned variants of the model. Together, this work provides the first unified computational theory of how perception and social cognition support contextual flexibility in real-time visual communication.