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

November 29, 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: they require a detailed understanding of how sensory information is encoded and how context guides what information is relevant to communicate.

Here we meet this challenge by providing a unified task paradigm and computational framework for investigating contextual flexibility in real-time visual communication tasks.

Participants (N=192) were paired in an online environment to play a drawing-based reference game. The *sketcher's* goal was to produce drawings so that the *viewer* could pick out a target object from a set of distractor objects.

On each trial, both participants were shown an array of the same four objects, but in different locations. One object was highlighted on the sketcher's display to designate it as the target.

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, participants achieved 99.7% accuracy while applying fewer strokes, using less ink, and spending less time ($ps < 0.001$) on their drawings than on close trials, where accuracy was still high (87.7%).

We hypothesized that humans succeed 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; (2) **social reasoning**, the ability to infer what information to draw to help the viewer to distinguish the target from distractors.

We instantiated these competencies in a context-sensitive sketcher model that combines a multimodal convnet visual encoder with a Bayesian model of social reasoning during communication.

Critically, this model outperformed a context-insensitive baseline that ignored the distractors.

Together, this work provides a unified model of how perception and social inference are integrated to support visual communication.

In the long run, investigating the computational basis of visual communicative flexibility may shed light on the sources of variation in pictorial style and the emergence of graphical conventions.

Suggested Reviewer 3: XXX