

Unconventional communication: making meaning across modalities

Arnav Verma¹, Lionel Wong², Matthew Caren²,
Robert D. Hawkins², Arvind Satyanarayan¹, Judith E. Fan²

¹Massachusetts Institute of Technology, ²Stanford University
{arnavv, arvindsatya}@mit.edu, {liowong, mcaren, rdhawkins, jefan}@stanford.edu

Keywords: communication; language; visual production; gesture; demonstration; music

Overview and Motivation

Humans are social beings, and intentional communication is the backbone of how people share what they are thinking, coordinate over long periods of time, and learn from each other’s ideas. People have finely attuned capacities for inferring each other’s mental states (Dennett, 2017), but intentional communication offers a much richer and more direct window into someone else’s thoughts than passive mind-reading alone.

A good portion of human communication takes place in well-developed natural languages—structured systems of conventionalized signs and meanings that allow people to externalize and recover arbitrary thoughts. Within cognitive science, in turn, much of the study of human communication has focused on hallmarks of established linguistic systems, such as the compositionality and systematicity of their vocabularies (Harris, 2025; Grice, 1975; Fodor, 1975). Extensive theoretical and empirical work has established that these hallmarks hold across the incredible range of modalities that people use for language. The grammar and vocabularies of well-developed natural languages can be encoded into sounds, visual symbols, tactile stimuli, and motor signs (Cohn & Schilperoord, 2022; Tversky, 2011). These signatures of human languages, and their emergence across modalities, have led many to suggest that people draw on specialized, modality-independent, and possibly evolved mechanisms for communicating in language (Pinker & Bloom, 1990).

Real-world communication, however, often seems to draw on a much more flexible and ad-hoc palette of cues and signals. People can repurpose seemingly any modality for this kind of ad-hoc communication: a blockade in a doorway to warn people of danger ahead (Jara-Ettinger & Schachner, 2024; Lopez-Brau & Jara-Ettinger, 2023); a freehand sketch to reference an unseen object (Fan et al., 2023); a frantic gesture to steer someone’s attention (Royka et al., 2021); a yelp for help (Mehr et al., 2021). Any given communicative context might freely mix modalities, and blend conventional language with ad-hoc communicative actions (like punctuating a speech with a gesture). But many of these more ad-hoc communicative actions do not obviously share the foundational hallmarks of conventionalized language. They are often not fully compositional, have far more uncertain and less clearly defined semantics, and can take place on much more extended timescales (like a dirty dish placed prominently in the center

of the kitchen table for someone to interpret later on). Ad-hoc communicative signals do not always develop into natural languages, and do not seem to require direct prior instruction or convention—people can produce and interpret one-off signals, without intending to standardize or use them again. Further, while unified natural languages systems can be implemented fully in a wide range of modalities, the ways that people use and interpret ad-hoc communicative signals often seem to draw much more closely on modality-specific affordances and constraints. We can interpret the blockade in the door the very first time we see it because of what we know about buildings and doorways, and it might be harder to send the same message as directly using an unplanned, non-linguistic set of sounds.

What determines how and when people make use of any given modality for ad-hoc communication? How do these ad-hoc actions and signals relate to someone’s existing knowledge of conventionalized language, their broader capacities for social reasoning, and their understanding of the unique affordances and constraints of that specific form? What cognitive mechanisms, in turn, allow people to integrate these varying communicative acts into flexible real-world communication—which freely mixes modalities, and moves between ad-hoc signals and conventionalized language? Existing work has developed theories that describe how conventionalized languages emerge from unstructured signals over time (Hawkins et al., 2023; Kirby et al., 2015). This workshop focuses, instead, on the inverse problem: how do people make flexible use of any given modality for ad-hoc and one-off communication, and how do they do that within the context of any more general communicative setting?

Our workshop will bring together discussants who have explored human communication across a wide range of distinct modalities, including conventional natural language, visual media, actions, sound, and environmental manipulations. They represent a diverse range of theories about how people make use of these modalities to communicate—some attribute meaning-making to innate, modality-specific semantics (Mehr et al., 2021), and others to much more general mechanisms for reasoning and inference (Jara-Ettinger, 2019). Together, however, our goal in this workshop is to spur discussion towards a unifying theoretical understanding of how people achieve flexible communication, across modality and degrees of convention. The resulting workshop will advance our understanding of the relationship between natural language and human communication more generally, explaining how we develop highly contextual and ad-hoc signals in real

time. It will also lay the groundwork for developing unifying formalisms that can be applied to understand and facilitate communication between humans and other agents, like effective, flexible multi-modal, and embodied human-robotic and human-computer interaction.

Approach and Schedule

This half-day workshop brings together researchers from cognitive science, linguistics, and philosophy which each perform empirical research across a broad range of different modalities. Our program will include 5 sessions with 15 minutes for invited talks, each followed by 15 minutes for audience questions and discussion. Each session will include a speaker who will answer three central questions: (1) How does the relevant modality's properties shape its use in communication, both in ad-hoc interactions and in conventionalized systems? (2) What is the relationship between conventional language and ad-hoc communication? and (3) How do people flexibly make use of any given modality for ad-hoc communication within the context of broader communicative settings?

Following each talk, the organizers will moderate a 5-minute question-and-answer session with the speaker, followed by a 10-minute discussion among the organizers, speakers, and audience members in small groups. We schedule the workshop as follows:

Opening Remarks

Session #1: Daniel W. Harris will speak on *natural language* communication.

Session #2: Neil Cohn will speak on *visual* communication.

Session #3: Julian Jara-Ettinger will speak on communication through *gesture and action*.

Session #4: Samuel Mehr will speak on communication through *sound and music*.

Session #5: Adena Schachner will speak on communication through *environment design*.

Closing Remarks

Speakers

Daniel W. Harris is an Associate Professor of Philosophy at Hunter College and the City University of New York Graduate Center. His research examines how theory of mind and language interact to enable people to tailor what they communicate and how they communicate it.

Neil Cohn is an Associate Professor at the Department of Communication and Cognition at Tilburg University. He studies visual languages, focusing on how meaning is structured and conveyed in comics and other visual media.

Julian Jara-Ettinger is an Associate Professor in Psychology at Yale University. His work investigates how people communicate using a rational theory of mind.

Samuel Mehr is a Senior Lecturer at the University of Auckland School of Psychology and an Associate Professor Adjunct at Yale University Child Study Center. He studies how the communicative semantics of music emerge through evolutionary pressures.

Adena Schachner is an Associate Professor of Psychology at the University of California, San Diego. Her research explores how people design and manipulate their environments to support communication.

Organizers

Arnav Verma is a PhD student at MIT who studies how people extract insight from and communicate through graphical media such as data visualizations.

Lionel Wong is a Postdoctoral Fellow at Stanford who investigates how people integrate information from natural language with their prior knowledge ad hoc for reasoning.

Matthew Caren is a PhD student at Stanford whose research explores human expression and communication through sound.

Robert D. Hawkins is an Assistant Professor at Stanford who studies how people flexibly communicate, collaborate, and coordinate in social interaction.

Arvind Satyanarayan is an Associate Professor at MIT whose work focuses on building interactive data visualization grammars and languages that support reasoning.

Judith E. Fan is an Assistant Professor at Stanford whose work investigates how people use physical tools and artifacts to support learning, communication, and problem solving.

References

- Cohn, N., & Schilperoord, J. (2022). Reimagining language. *Cognitive Science*, 46(7), e13174.
- Dennett, D. C. (2017). *Brainstorms: Philosophical essays on mind and psychology*. MIT press.
- Fan, J. E., Bainbridge, W. A., Chamberlain, R., & Wammes, J. D. (2023). Drawing as a versatile cognitive tool. *Nature Reviews Psychology*, 2(9), 556–568.
- Fodor, J. A. (1975). *The language of thought* (Vol. 5). Harvard university press.
- Grice, H. P. (1975). Logic and conversation. In *Speech acts* (pp. 41–58). Brill.
- Harris, D. W. (2025). Gricean communication, natural language, and human evolution. *Evolutionary Pragmatics*.
- Hawkins, R. D., Franke, M., Frank, M. C., Goldberg, A. E., Smith, K., Griffiths, T. L., & Goodman, N. D. (2023). From partners to populations: A hierarchical bayesian account of coordination and convention. *Psychological Review*, 130(4), 977.
- Jara-Ettinger, J. (2019). Theory of mind as inverse reinforcement learning. *Current Opinion in Behavioral Sciences*, 29, 105–110.
- Jara-Ettinger, J., & Schachner, A. (2024). Traces of our past: The social representation of the physical world. *Current Directions in Psychological Science*, 33(5), 334–340.
- Kirby, S., Tamariz, M., Cornish, H., & Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition*, 141, 87–102.
- Lopez-Brau, M., & Jara-Ettinger, J. (2023). People can use the placement of objects to infer communicative goals. *Cognition*, 239, 105524.

- Mehr, S. A., Krasnow, M. M., Bryant, G. A., & Hagen, E. H. (2021). Origins of music in credible signaling. *Behavioral and Brain Sciences*, *44*, e60.
- Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Behavioral and brain sciences*, *13*(4), 707–727.
- Royka, A. L., Schouwstra, M., Kirby, S., & Jara-Ettinger, J. (2021). I know you know i'm signaling: Novel gestures are designed to guide observers' inferences about communicative goals. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 43).
- Tversky, B. (2011). Visualizing thought. *Topics in Cognitive Science*, *3*(3), 499-535. Retrieved from <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1756-8765.2010.01113.x> doi: <https://doi.org/10.1111/j.1756-8765.2010.01113.x>