Predicting Pragmatic Reasoning in Language Games

Michael C. Frank* and Noah D. Goodman

One of the most astonishing features of human language is its ability to convey information efficiently in context. Each utterance need not carry every detail; instead, listeners can infer speakers' intended meanings by assuming utterances convey only relevant information. These communicative inferences rely on the shared assumption that speakers are informative, but not more so than is necessary given the communicators' common knowledge and the task at hand. Many theories provide high-level accounts of these kinds of inferences (*1–3*), yet, perhaps because of the difficulty of formalizing notions like "informativeness" or "common knowledge," there have been few successes in making quantitative predictions about pragmatic inference in context.

We addressed this issue by studying simple referential communication games, like those described by Wittgenstein (*4*). Participants see a set of objects and are asked to bet which one is being referred to by a particular word. We modeled human behavior by assuming that a listener can use Bayesian inference to recover a speaker's intended referent r_S in context C, given that the speaker uttered word w:

$$P(r_s|w,C) = \frac{P(w|r_s,C)P(r_s)}{\sum_{r' \in C} P(w|r',C)P(r')} \quad (1)$$

This expression is the product of three terms: the prior probability $P(r_S)$ that an object would be referred to; the likelihood $P(w|r_S,C)$ that the speaker would utter a particular word to refer to the object; and the normalizing constant, a sum of these terms computed for all referents in the context.

We defined the prior probability of referring to an object as its contextual salience. This term picks out not just perceptually but also socially and conversationally salient objects, capturing the common knowledge that speaker and listener share, as it affects the communication game. Because there is no a priori method for computing this sort of salience, we instead measured it empirically (*5*).

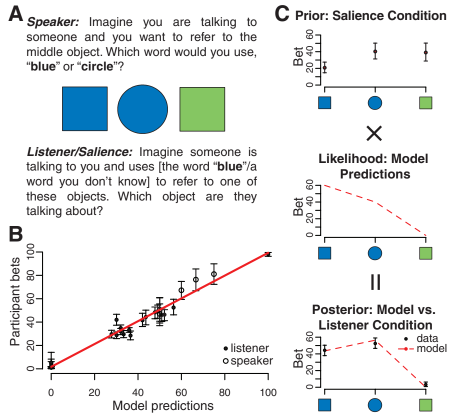

Fig. 1. (**A**) An example stimulus from our experiment, with instructions for speaker, listener, and salience conditions. (**B**) Human bets on the probability of a choosing a term (speaker condition, $N = 206$) or referring to an object (listener condition, $N = 263$), plotted by model predictions. Points represent mean bets for particular terms and objects for each context type. The red line shows the best linear fit to all data. (**C**) An example calculation in our model for the context type shown in (A). Empirical data from the salience condition constitute the prior term, $N = 20$ (top); this is multiplied by the model-derived likelihood term (middle). The resulting posterior model predictions (normalization step not shown) are plotted alongside human data from the listener condition, $N = 24$ (bottom). All error bars show 95% confidence intervals.

The likelihood term in our model is defined by the assumption that speakers choose words to be informative in context. We quantified the informativeness of a word by its surprisal, an information-theoretic measure of how much it reduces uncertainty about the referent. By assuming a rational actor model of the speaker, with utility defined in terms of surprisal, we can derive the regularity that speakers should choose words proportional to their specificity (*6*, *7*):

$$P(w|r_s,C) = \frac{|w|^{-1}}{\sum_{w' \in W} |w'|^{-1}} \quad (2)$$

where $|w|$ indicates the number of objects to which word w could apply and W indicates the set of words that apply to the speaker's intended referent.

In our experiment, three groups of participants each saw communicative contexts consisting of sets of objects varying on two dimensions (Fig. 1A). We systematically varied the distribution of features on these dimensions. To minimize the effects of particular configurations or features, we randomized all other aspects of the objects for each participant. The first group (speaker condition) bet on which word a speaker would use to describe a particular object, testing the likelihood portion of our model. The second group (salience condition) was told that a speaker had used an unknown word to refer to one of the objects and was asked to bet which object was being talked about, providing an empirical measure of the prior in our model. The third group (listener condition) was told that a speaker had used a single word (e.g., "blue") and again asked to bet on objects, testing the posterior predictions of our model.

Mean bets in the speaker condition were highly correlated with our model's predictions for informative speakers ($r = 0.98$, $P < 0.001$; Fig. 1B, open circles). Judgments in the salience and listener conditions were not themselves correlated with one another ($r = 0.19$, $P = 0.40$), but when salience and informativeness terms were combined via our model, the result was highly correlated with listener judgments ($r = 0.99$, $P < 0.0001$, Fig. 1B, solid circles). This correlation remained highly significant when predictions of 0 and 100 were removed ($r = 0.87$, $P < 0.0001$). Figure 1C shows model calculations for one arrangement of objects.

Our simple model synthesizes and extends work on human communication from a number of different traditions, including early disambiguation models (*8*), game-theoretic signaling models (*9*), and systems for generating referring expressions (*10*). The combination of an information-theoretic definition of "informativeness" along with empirical measurements of common knowledge enables us to capture some of the richness of human pragmatic inference in context.

References and Notes

1. H. Grice, in *Syntax and Semantics*, P. Cole, J. Morgan, Eds. (Academic Press, New York, 1975), vol. 3, pp. 41–58.
2. D. Sperber, D. Wilson, *Relevance: Communication and Cognition* (Harvard Univ. Press, Cambridge, MA, 1986).
3. H. Clark, *Using Language* (Cambridge Univ. Press, Cambridge, 1996).
4. L. Wittgenstein, *Philosophical Investigations* (Blackwell, Oxford, 1953).
5. H. Clark, R. Schreuder, S. Buttrick, *J. Verbal Learn. Verbal Behav.* **22**, 245 (1983).
6. Materials and methods are available as supplementary materials on *Science* Online.
7. F. Xu, J. B. Tenenbaum, *Psychol. Rev.* **114**, 245 (2007).
8. S. Rosenberg, B. D. Cohen, *Science* **145**, 1201 (1964).
9. A. Benz, G. Jäger, R. Van Rooij, Eds., *Game Theory and Pragmatics* (Palgrave Macmillan, Hampshire, UK, 2005).
10. R. Dale, E. Reiter, *Cogn. Sci.* **19**, 233 (1995).

Supplementary Materials
www.sciencemag.org/cgi/content/full/336/6084/998/DC1
Materials and Methods
Supplementary Text

3 January 2012; accepted 10 April 2012
10.1126/science.1218633

Department of Psychology, Stanford University, Stanford, CA 94305, USA.

*To whom correspondence should be addressed. E-mail: mcfrank@stanford.edu