

**Testing a rational account of pragmatic reasoning: The case of
spatial relations**

Cognitive Science Honor's Thesis by:

Elizabeth Kon

Advisor:

Terry Regier

Second Reader:

Thomas Griffiths

Abstract

Frank and Goodman (2012) proposed that listeners understand ambiguous utterances by rationally combining evidence about word meaning and the salience of particular objects in context. They found that a Bayesian statistical model using this information provided a near-perfect account of their empirical data. However, their test of the model was based on communication about simple geometrical objects that varied along only three dimensions. Here, we ask whether their proposal extends to the richer and more complex domain of spatial relations. We find that it does. While the results are not as strong as in their original study, they nonetheless demonstrate that simple formal accounts of communication may capture important aspects of pragmatic inference.

Language, thought, and communication

A growing trend in cognitive science views language through the lens of its function: as a vehicle for *informative, efficient communication* (e.g. Piantadosi et al., 2011, Fedzechkina et al., 2012). One such line of work has argued that systems of word meanings from various semantic domains in the world's languages tend to support highly informative communication (Regier et al., 2007; Baddeley & Attewell, 2009; Kemp & Regier, 2012; Khetarpal et al., 2013). For example, in the field of kinship, Kemp and Regier (2012) found that cross-linguistically, kinship categories optimize the trade-off between simplicity and informativeness, meaning that the system is easy to learn/use, but still allows for the communication of a lot of information. To this end, Kemp and Regier found that kinship categories are unlikely to be disjunctive (e.g. "either a woman or younger") and existing systems more optimally divide the space than other theoretically possible systems. However, word meanings necessarily leave much information unspecified; it would be inefficient and impossible to have a unique word for every socket

wrench or pencil. Therefore, ambiguity is an important aspect of this system. Piantadosi et al. demonstrated that a fully unambiguous communication system, while informative, would not be efficient because information would be redundantly encoded in both linguistic form and context (Piantadosi et al., 2012). Additionally, Levy (2007) has shown that within a parallel processing information theoretic model of syntactic processing in which the difficulty of a word is proportional to its surprisal in context, ambiguity can actually facilitate linguistic processing. It accomplishes this because multiple analyses (based on different interpretations of ambiguous words) conspire to ease the processing of a word. Levy's model captures the empirical results of Traxler et al. (1998), van Gompel et al. (2001), and van Gompel et al. (2005) who found that ambiguous left attachments to a complex noun phrase are read faster when they are not resolved at the attachment level. Because ambiguity is built into the system of word meanings, the use of words must be supplemented by pragmatic reasoning to allow speaker and listener to communicate effectively. What principles govern this pragmatic reasoning?

The study of pragmatics is often traced back to Paul Grice who argued that when conversing, people operate following a cooperative principal in which each person obeys three maxims: the maxim of quality (they say what they believe to be true), the maxim of quantity (they say exactly as much as necessary to convey the information), and the maxim of relevance (they say things that have bearing on the conversation being had). Because the listener assumes that their conversational partner is operating in accordance with these maxims, they can extract more meaning from utterances than is conveyed by the semantics alone (Grice, 1975). From there, the ideas of investigating language in context took off, and linguists began to study not only what was said, but how listeners were likely to interpret it based on what had previously been said in

the conversation, what was happening in the physical world around the conversation, and the known biases of the interlocutors (e.g. Clark, 1996; Sperber & Wilson, 1996).

Additionally, a significant amount of research has been focused on examining the function of common ground in allowing listeners to better understand speakers. That is, researchers have looked at how well conversational partners can capitalize on their shared knowledge. One such study, Hanna et al. (2003), ran a visual world experiment in which some items were seen by both the speaker and the listener (common ground) and other objects were seen only by the listener (privileged ground). They found that listeners were always more likely to look at items in the common ground over the privileged ground, but the items in the privileged ground did interfere with their processing. Many other studies have supported the importance of integrating the knowledge of a conversational partner when understanding what they are saying (e.g. Breheny et al., 2013; Clark et al., 1983; Heller et al., 2008; Liszkowski et al., 2008).

Recently, several research teams have been interested in formalizing processes of pragmatic inference. The earliest models focused on speech acts; for example, the plan inference model of speech act interpretation (Allen, 1995), and the cue-based models of speech act interpretation (e.g. Stolcke et al., 2000; Shriberg et al., 1998). More recently, people have extended models into the realm of referring expressions; for example, the iterated best response model is a game-theoretic model which takes into account the recursive understanding of the speaker's beliefs about the listener's beliefs and the listener's beliefs about the speaker's beliefs etc (Franke, 2009; Benz et al., 2005; and the newest version now called the iterated quantal response model: Degen et al., 2013).

Frank & Goodman (2012) proposed an elegant account of pragmatic reasoning in language use: the rational speech act model. They argued that in conversation, listeners determine the

object to which a speaker is referring by rationally combining two sorts of evidence: one concerning how well the speaker's utterance fits each potential referent, and the other concerning how salient each potential referent is in context. Their study presented evidence that listeners combine these two sources of evidence, in accord with Bayes' rule, to use in interpreting the speaker's intention. It has been noted that there are several similarities between the rational speech act model and Kemp and Regier's model of efficient communication mentioned earlier in that both take an information-theoretic approach to address the concept of informativeness (Levinson, 2012). One limitation of Frank and Goodman's study is that it was based on communication about a very simple and cleanly circumscribed semantic domain, and it is not yet known whether similar results would be obtained in a more complex domain.

We seek to answer that question here by replicating their study in the context of communication about *spatial relations*: the ways in which one object is situated spatially relative to another object. Languages differ substantially in the means by which they partition the spatial domain into semantic categories, and these categories sometimes involve relatively subtle features such as attachment by spiking, or being astraddle, in addition to (from a Western viewpoint) more obvious features such as containment and support (Levinson et al., 2003). Spatial relationships have been broadly studied within the cognitive science (e.g. Munnich et al., 2001; Khetarpal et al., 2013; Knauff et al., 2005; Carstensen & Regier, 2013; McNamara, 1986). Thus, the domain of spatial relations is rich enough to allow a test of Frank & Goodman's (2012) proposal in a semantically complex domain.

We first review Frank and Goodman's study, on which ours is based. We then present our study, which tests their proposal in the spatial domain. To preview our results, we find that their account does predict pragmatic reasoning in the spatial domain, but does not do so as cleanly as

in their original study in a simpler domain. We conclude that pragmatic reasoning in more complex domains is substantially but not fully accounted for by their proposal as it stands, and consider possible interpretations of this finding.

Frank & Goodman (2012)

How does a listener interpret a speaker's utterance in context? Imagine that a speaker wishes to refer to a specific referent r_s , which along with several other possible referents make up the physical context C , and that the speaker has produced a word w to convey this to a listener. Frank & Goodman (2012) proposed that in such situations, the listener determines the speaker's intended referent through Bayes' rule:¹

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

Here, the posterior probability $P(r_s | w, C)$ represents the listener's subjective degree of belief that the speaker's intended referent is r_s , given word w and context C . This quantity is proportional to the product of two terms: 1. the likelihood $P(w | r_s, C)$ of the speaker using word w given that the intended referent was r_s in context C , and 2. the prior probability $P(r_s | C)$ that a word in context C would refer to r_s , without any specification of what that word is. The denominator of Equation (1) is a normalizing constant to give us a probability that the intended referent is r_s when all other possible referents are taken into consideration.

Frank & Goodman (2012) assumed that speakers choose words to be maximally specific—that is, that speakers select the term that picks out the smallest set of possible referents in a given context (cf. Xu & Tenenbaum, 2007). Accordingly, they modeled the likelihood $P(w | r_s, C)$ as the size principle:

¹ Our notation differs slightly from that of Frank and Goodman (2012), but there is no difference in intended meaning.

² We describe their conditions briefly here, and provide concrete examples with spatial stimuli below,

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

where w is the selected word, $|w|$ is the number of objects in the extension of w , and W is the set of all labels that could be validly applied to the intended referent r_s .

To test their model, Frank and Goodman conducted an experiment with three conditions: one to assess each of the model’s three components.² In all three conditions, participants viewed a communicative context C consisting of three simple geometrical objects that could vary in shape, color, and texture. For example, a context might contain a solid blue square, a solid blue circle, and a solid green circle. In the *speaker* condition, one of the objects in the context was highlighted as the intended referent (e.g. the blue circle), and participants were asked to bet on which word (e.g. “blue”, “circle”) they would use to describe that object in that context; this provides an empirical measure of the likelihood $P(w|r_s, C)$. In the *salience* condition, no object was highlighted—instead, participants were told that a speaker had used an unknown word to refer to one of the objects shown in the context, and they were asked to bet on which object was intended; this provides an empirical measure of the prior $P(r_s|C)$. Finally, in the *listener* condition, participants again saw three objects in context without any object highlighted, but this time were told that a speaker had used a single word (e.g. “blue”) to refer to one of the objects, and were asked to bet on which object the speaker intended; this provides an empirical measure of the posterior $P(r_s|w, C)$. Frank and Goodman found that mean bets in the speaker condition were very highly correlated with their model likelihood (Equation 2), and that mean bets in the listener condition were very highly correlated with their model posterior probability (Equation

² We describe their conditions briefly here, and provide concrete examples with spatial stimuli below, when we present our variant of their experiment.

1). They concluded that this simple model captures “some of the richness of human pragmatic inference in context.”

We wished to test whether Frank and Goodman’s results generalize to the more complex domain of spatial relations. To that end, we followed their formalization, their experimental design, and their analysis, changing only the character of the stimuli and the words that refer to them, as described below.

Our study

In our experiment, we replaced Frank and Goodman’s simple geometric stimuli with line drawings that depict spatial relations. We chose the domain of spatial relations because several cross-linguistic studies have shown that this domain contains underlying complexity not present in the geometric stimuli used in the original study (e.g. Bowerman & Pederson, 1992; Levinson et al., 2003; Munnich et al., 2001; Khetarpal et al., 2013). Our stimuli were taken from the Topological Relations Picture Series (TRPS; Bowerman & Pederson, 1992), a set of 71 line drawings depicting a variety of spatial relations. Each line drawing shows an orange figure object located relative to a black background object. Figure 1 shows a sample of 10 scenes from the TRPS, categorized according to the spatial naming systems of two languages.

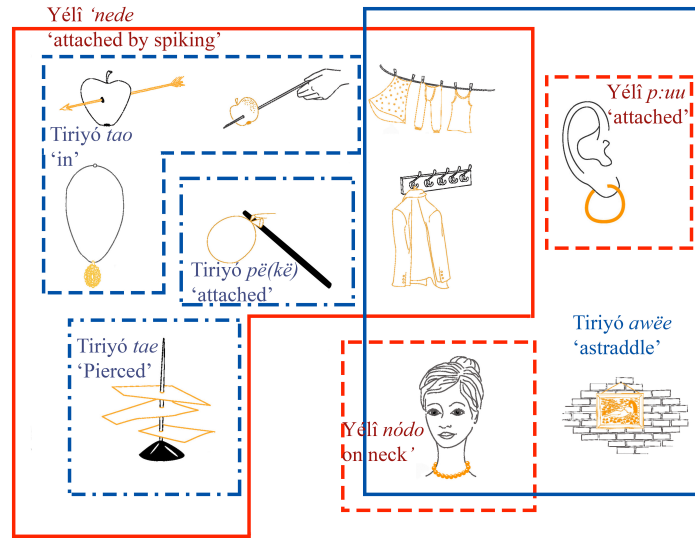


Figure 1: Ten spatial scenes from the TRPS, categorized in the languages Tiriyo and Yeli-Dnye. Adapted from Levinson et al., (2003).

The TRPS has been widely used in cross-linguistic studies of spatial language (e.g. Bowerman & Pederson, 1992; Levinson et al., 2003; Khetarpal et al., 2013; Regier et al., 2013), and it represents a broad, rich, and finely-detailed range of different spatial relations. We investigate pragmatic reasoning about reference in the domain of such scenes, using the spatial terms of English (e.g. “in,” “around,” etc.).

Methods

Participants

A total of 1,427 participants from the U.S. took part in our experiment online through Amazon Mechanical Turk. These participants completed a total of 1,605 trials across all conditions (described below), 447 of which trials were excluded from our analysis because the participant

either failed to follow instructions or completed more than one trial, in which case subsequent trials were discarded. Because each trial in every condition was completed by a unique participant, the number of participants in each condition is equal to the number of trials.

Materials

Communicative contexts were constructed as triads of TRPS scenes presented side by side; an example is shown in Figure 2 below. We divided the 71 scenes of the TRPS into 23 unique triad sets, such that no scene appeared in more than one triad, and we excluded the two remaining scenes (TRPS scenes 2 and 46). Because the goal of the study is to investigate reasoning under ambiguous reference, each triad was formed with the requirement that English spatial terms should be ambiguous when used in the context of the triad. Specifically, every member of the triad shared at least one English spatial term that could describe that scene (collection described below) with another member of the triad.

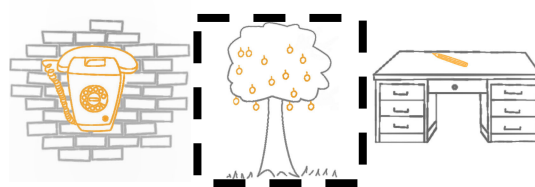
Determination of spatial terms for scenes

In a separate experiment, 45 UC Berkeley undergraduates, all native English speakers, viewed each TRPS scene and answered the question “Where is the [figure object]?” by completing a fill-in-the-blank sentence that specified figure and ground but not the spatial relationship between them, for instance: “The cup _____ the table.” Responses were trimmed to standardize tense and remove non-spatial words (e.g. “is”). To ensure that spatial terms were all of similar complexity, we only included responses with two or fewer spatial morphemes as valid spatial term options. A spatial morpheme is a meaningful unit that conveys information about the spatial relationship between figure and ground objects, so *around* would have one spatial morpheme, as would *next to*, but *hanging on* has two. This procedure resulted in each TRPS scene receiving at least two spatial term labels; many received more.

Design and procedure

Our design directly paralleled that of Frank and Goodman (2012). There were three conditions, corresponding to the three elements of their model. The speaker condition empirically measured the likelihood $P(w|r_s, C)$; the salience condition empirically measured the prior $P(r_s|C)$; and the listener condition empirically measured the posterior $P(r_s|w, C)$. In all conditions, participants viewed triads of spatial scenes (contexts) and answered questions about them. Figure 2 shows an example trial, with instructions from each of the three conditions.

Speaker (likelihood): Imagine you are talking to someone and you want to refer to the selected scene and distinguish it from the other two scenes. Which word would you use, “**on**” or “**hanging from**”? Estimate the probability that you would use each word as a percent (responses must add to 100).



Salience (prior) / Listener (posterior): [scene selection highlight not shown] Imagine someone is talking to you and uses [a word you don’t know/the word “**on**”] to refer to one of these scenes. Which scene are they talking about? Estimate the probability that they are talking about each of the scenes as a percent (responses must add to 100).

Figure 2: An example trial, with instructions from the speaker, salience, and listener conditions. For each of the three conditions, we specify below any elements of procedure not already specified.

Speaker (likelihood). Participants viewed a triad of spatial scenes, one of which was selected as the intended referent. The selected scene (the intended referent) was always indicated by a dotted black square around it. Participants were given a list of all valid spatial terms (identified by the earlier experiment described above) that could be applied to the intended referent, and were instructed to estimate the probability that they would use each term in the set to refer to the selected scene, in the context of that triad of scenes. The instructions specified that these probability estimates should add to 100, and this requirement also served as a comprehension check; trials in which participants' estimates did not sum to 100 were discarded and re-run on new participants. Any of the three scenes within a triad could be the selected referent, yielding 3 (scenes per triad) \times 23 (triads) = 69 unique trial types (a trial type is a triad with a particular scene selected, as in Figure 2). Scene order was fully counterbalanced within these trial types for a total of 6 orders \times 69 trial types = 414 trials in this condition.

Salience (prior). Each participant was shown a triad of spatial scenes without any scene selected, and instructed to imagine that someone had used a word that the participant did not know to refer to one of the scenes in the triad. They were asked to estimate how likely it was that the speaker was referring to each scene, such that their estimates summed to 100. As in the speaker (likelihood) condition, trials in which the participant failed to follow this instruction were discarded and re-run with new participants. Scene order was fully counterbalanced within the triad sets for a total of 6 orders \times 23 triads = 138 trials in this condition.

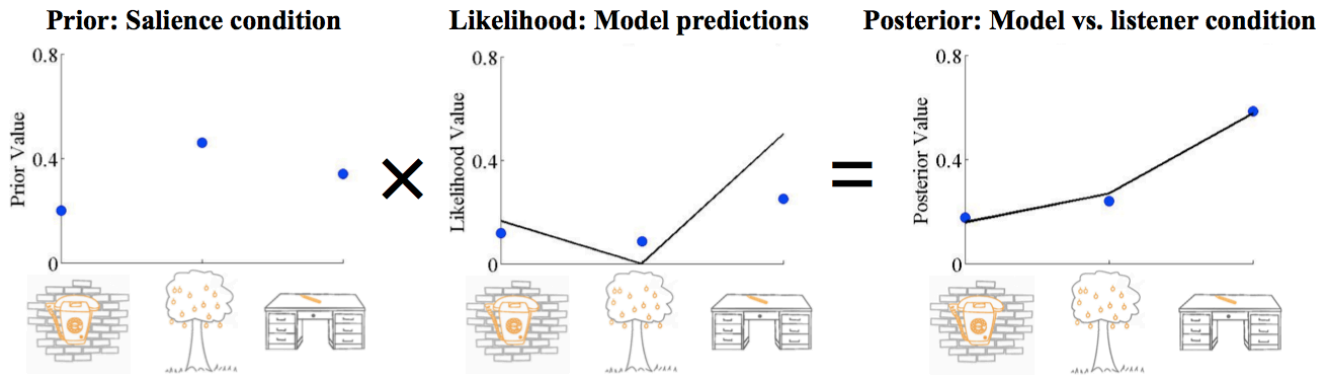


Figure 3: Example calculation in the model for the scene triad shown in Figure 2. Blue dots indicate participants' average responses; black lines indicate model predictions.

Listener (posterior). Each participant was shown a triad of spatial scenes without any scene selected, together with an English spatial term (e.g. “on”) that could be validly applied to at least one scene in the triad. They were asked which scenes in the triad a speaker might be talking about when using that label. Specifically, participants judged how likely it was that each scene was the speaker’s intended referent given that spatial term, and entered their judgments as percentages summing to 100. As in the other two conditions, participants whose estimates did not sum to 100 were excluded and the trials re-run. Each of the 23 triads was paired with all possible labels for scenes in that triad, yielding 202 unique trial types. Order within these trial types was pseudo-randomly counterbalanced such that each unique trial type was presented in three of the six possible scene orders, yielding 606 trials in total in this condition.

Analysis and results

Our analyses followed those of Frank & Goodman (2012). We first tested the model’s assumption of speaker informativeness. We then tested whether saliency (the prior) predicts responses in the listener condition—to see whether this one source of evidence by itself suffices

to explain listeners' inferences. Finally, we assessed the combination of evidence through Bayes' rule, by comparing the model posterior to empirical responses in our listener condition (to which this quantity is intended to correspond) together with a follow-up analysis. Figure 3 illustrates model calculations alongside empirical results for one sample triad of scenes.

Testing the assumption of speaker informativeness

The model likelihood (Equation 2) is based on the assumption that speakers choose words to be maximally informative in context—that is, so that the word chosen will pick out the smallest set of referents possible in a given context. We tested this assumption by comparing empirical data in the speaker (likelihood) condition with the model likelihood term obtained through Equation 2. We found a significant correlation between average empirical likelihoods and model predictions ($r = .36, p < .0001$). This result suggests that the model likelihood reasonably approximates speakers' word choice in context, and that speakers do appear to choose their words informatively.

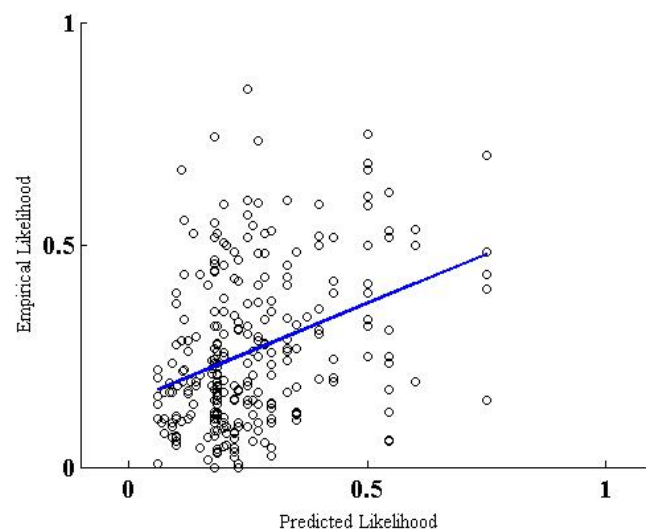


Figure 4: Correlation between model likelihood (size principal) and participants' judgments about word they would use in context

Does salience alone predict listener's inferences?

It is conceivable that listeners might base their judgments of speakers' intentions solely on the salience of particular objects, without reference to how well a given word fits each referent. To test this, we compared empirical data from the salience (prior) condition to empirical data from the listener (posterior) condition. We found no significant correlation ($r = .06, p = .17$). This means that if the Bayesian model's posterior successfully predicts data from the listener (posterior) condition, that success cannot be due only to the prior, independent of likelihood.

Testing the Bayesian model

Finally, we tested the central claim: that listeners infer speakers' intentions through Bayesian combination of evidence. We combined the empirical prior (from the salience condition) and model likelihood (from Equation 1), to obtain the model's predicted posterior—and compared it to the empirical posterior (listener condition). We found a significant correlation ($r = .70, p < .0001$; see Figure 5).

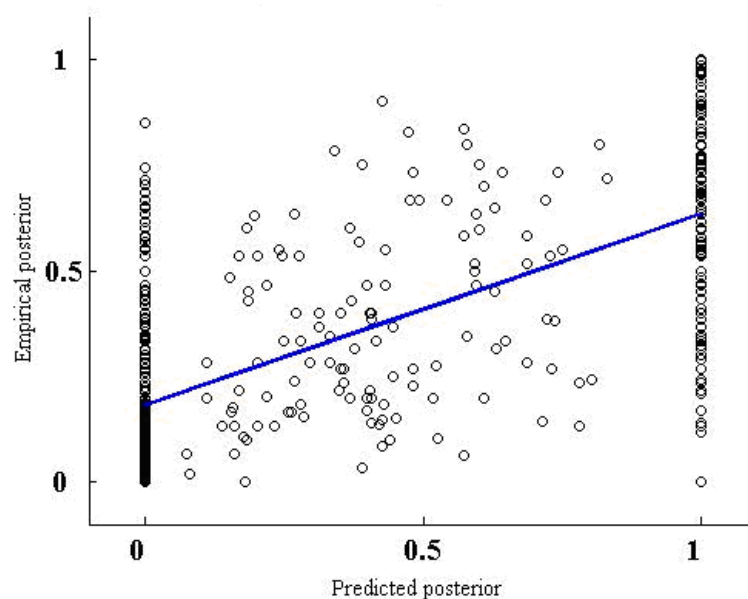


Figure 5: Correlation between model prediction and participants' judgments about which spatial scene the speaker intended, given a speaker's spatial term used in context

This correlation remains significant when predicted posterior values of one and zero are excluded ($r = .36, p < .0001$). We also explored another way to obtain a predicted posterior via Bayes' rule: using the empirically determined likelihood (speaker condition), rather than the model likelihood, such that Bayes' rule is now used to combine two empirically determined sources of evidence. In this case, the correlation with the empirical posterior (listener condition) was again significant ($r = .70, p < .0001$). Thus, Frank & Goodman's (2012) Bayesian account does seem to capture listeners' inference about speakers' intentions under conditions of referential uncertainty, in a complex and semantically rich domain.

Given these results, and given that we have also found that any success of the Bayesian model cannot be attributed solely to the prior, we sought to understand whether the model's success could be attributed solely to the likelihood instead. To that end, we obtained predictions of listeners' judgments using Equation 1 again, but this time assuming a uniform prior ($P(r_s|C) = \frac{1}{3}, \forall r \in C$), and using the model likelihood of Equation 2. We found that the correlation between this uniform-prior-based model prediction and listener judgments was high ($r = .70, p < .0001$)—in fact, it was as high as the correlation we obtained when combining the model likelihood with the empirical prior (salience condition). Additionally, we repeated this uniform prior analysis using the empirical likelihood (speaker condition) rather than the model likelihood, and obtained very similar results ($r = .71, p < .0001$). Thus, it appears that this empirically-based prior adds nothing to the predictive power of the model, and the real predictive component is the likelihood. Table 1 summarizes the results of all our analyses, together with analogous analyses by Frank and Goodman (2012).

Frequency Analyses

In their discussion of the derivation of the model, Frank and Goodman note that while they assume in their case that cost is constant, in other situations the cost could be affected by other factors including utterance length, and frequency. Our utterances were limited to one word each, so this does not apply, but we did do a frequency analysis. We collected frequencies from the Corpus of Contemporary American English, which includes data from many genres of English (spoken, academic, fiction, magazine, and newspaper) from the years 1990-2012 (Davies, 2008). We also found the relative proportions of the literal spatial uses of our TRMs and found that correlations of the model likelihood and the empirical likelihood became weaker when weighted by the relative frequencies of these literal TRM frequencies making it no longer significant ($r=-.1107$, $p=.10$). When weighted by the relative proportions of all uses of the TRMs in COCA, we found that the correlation between the model and empirical likelihoods was again not significant ($r=-.0980$, $p=.11$).

Correlation	Present	Frank & Goodman
Likelihood: model vs. empirical	0.36*	0.98*
Empirical prior vs. empirical posterior	0.06	0.19
Model vs. empirical posterior	0.70*	0.99*
Bayes vs. empirical posterior	0.70*	—
Model with uniform prior vs. empirical posterior	0.70*	—
Bayes with uniform prior vs. empirical posterior	0.71*	—
Likelihood: model weighted by literal TRM frequencies vs. empirical	0.11	—
Likelihood: model weighted by total TRM frequencies vs. empirical	-0.10	—

Table 1: Pearson correlations in the present study compared with those of Frank & Goodman (2012).

Discussion and conclusions

We have shown that Frank and Goodman's formalization of pragmatic inference in conversation extends to the domain of spatial relations—a more complex and naturalistic domain than that of simple geometric objects, in which they originally assessed their proposal. This suggests that their ideas may extend to more complex and richer semantic domains.

However, our results also suggest caution in at least two respects. First, the correlations between the model prediction and our listener judgment data, while significant, are substantially weaker than those of Frank & Goodman (2012), which were remarkably strong. Frank and Goodman explicitly anticipated that other factors such as word length and frequency – which they provisionally assumed would not be relevant in their initial study – may be relevant more generally; a natural question is whether such factors account for the difference in model fit between our study and theirs. Second, we have seen that the model's success with our data is attributable entirely to the likelihood, and not at all to the prior. It is possible that our empirical prior is, for whatever reason, a flawed measure of the contextual salience of particular objects. This possibility cannot be assumed, but it also cannot be ruled out. Future research can usefully focus on other means of assessing contextual salience, to help resolve this issue.

One possible source for our lower correlations is that the extent based likelihood model is not sufficient to explain word choice in a domain with graded category membership. In Frank & Goodman's study, a label either could or could not be applied to an object and it applied equally well to all relevant referents. However, in our study, labels could be applied in varying degrees. For example, both the picture of a cup on top of a table and gum stuck under a table could be labeled 'on'; however, the cup picture is a better example of the relation, and as such might be labeled 'on' even in the presence of less good 'on's' in the context. This is essentially the same

problem addressed in Gatt et al. (2013), who argue for a trade-off between preference (using descriptors that are the most salient) and discriminatory power (using descriptors that refer to the fewest possible referents, the size principal) when selecting referring expressions. They found that discriminatory power plays a relatively unimportant role in word choice. This suggests that incorporating a measure of preference, which in our case would be a measure of the goodness of category membership of each scene for each spatial relationship, would be beneficial when considering the speaker's task of choosing referring expressions. Future research could investigate whether incorporating this gradedness of category membership into the likelihood term improves the model fit.

These caveats notwithstanding, our results extend Frank & Goodman's (2012) account to a richer and more complex semantic domain, helping to support their conclusion that simple, formal accounts of communication may capture important aspects of pragmatic inference.

Acknowledgments

We thank Michael Frank, Noah Goodman, Yang Xu, and Joshua Abbott for helpful comments.

This work was supported by NSF under grant SBE-1041707, the Spatial Intelligence and Learning Center (SILC), and under NSF Graduate Research Fellowship grant DGE 1106400.

References

- Allen, J. (1995). *Natural language understanding*. Menlo Park, CA: Benjamin Cummings.
- Baddeley, R., & Attewell, D. (2009). The relationship between language and the environment: Information theory shows why we have only three lightness terms. *Psychological Science, 20*, 1100-1107.
- Benz, A., Jäger, G., & van Rooij, R. (2005). *Game theory and pragmatics*. Palgrave Macmillan.
- Bergen, L., & Grodner, D. (2012). Speaker knowledge influences the comprehension of pragmatic inferences. *Journal of Experimental Psychology: Learning, Memory and Cognition, 38*, 1450-1460
- Bowerman, M., & Pederson, E. (1992). Cross-linguistic studies of spatial semantic organization. In *Annual Report of the Max Planck Institute for Psycholinguistics 1992*, 53-56.
- Breheny, R., Ferguson, H., Katsos, N. (2013). Taking the epistemic step: Toward a model of on-line access to conversational implicatures. *Cognition, 126*, 423-440.
- Carstensen, A., & Regier, T. (2013). Individuals recapitulate the proposed evolutionary development of spatial lexicons. In M. Knauff et al., (Eds.), *Proceedings of the 35th Annual Meeting of the Cognitive Science Society*.
- Clark, H. (1996). *Using language*. Cambridge University Press, 9780521567459
- Clark, H., Schreuder, R., & Buttrick, S. (1983). Common ground and the understanding of demonstrative reference. *Journal of Verbal Learning and Verbal Behavior, 22*, 245-258.
- Davies, Mark. (2008-) *The Corpus of Contemporary American English: 450 million words, 1990-present*. Available online at <http://corpus.byu.edu/coca/>.

- Degen, J., Franke, M., & Jäger, G. (2013). Cost-based pragmatic inference about referential expressions. . In M. Knauff et al., (Eds.), *Proceedings of the 35th Annual Meeting of the Cognitive Science Society*.
- Fedzechkina, M., Jaeger, T., & Newport, E. (2012). Language learners restructure their input to facilitate efficient communication. *PNAS*, *109*, 17897–17902.
- Frank, M., & Goodman, N. (2012). Predicting pragmatic reasoning in language games. *Science*, *336*, 998.
- Franke, M. (2009), Signal to act: Game theory in pragmatics, Ph.D. thesis, Institute for Logic, Language and Computation, University of Amsterdam
- Garrod, S., & Pickering, M. (2004). Why is conversation so easy?. *Trends in Cognitive Science*, *8*, 8-11.
- Gatt, A., Krahmer, E., van Gompel, R., & van Deemter, K. (2013). Production of referring expressions: Preference trumps discrimination. In M. Knauff et al., (Eds.), *Proceedings of the 35th Annual Meeting of the Cognitive Science Society*.
- Goodman, N., & Stuhlmüller, A. (2013). Knowledge and implicature: Modeling language understanding as social cognition. *Topics in Cognitive Science*, *5*, 173-184.
- Grice, P. (1975). Logic and conversation. *Syntax and Semantics*, *3*, 41-58.
- Hanna, J., Tanenhaus, M., & Trueswell, J. (2003). The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory and Language*, *49*, 43-61.
- Heller, D., Grodner, D., & Tanenhaus, M. (2008). The role of perspective in identifying domains of reference. *Cognition*, *108*, 831-836.

- Kemp, C., & Regier, T. (2012). Kinship categories across languages reflect general communicative principles. *Science*, 336, 1049-1054.
- Khetarpal, N., Neveu, G., Majid, A., Michael, L., & Regier, T. (2013). Spatial terms across languages support near-optimal communication: Evidence from Peruvian Amazonia and computational analyses. In M. Knauff et al., (Eds.), *Proceedings of the 35th Annual Meeting of the Cognitive Science Society*.
- Knauff, M., Rauh, R., & Renz, J. (2005). A cognitive assessment of topological spatial relations: Results from an empirical investigation. *Lecture Notes in Computer Science*, 1329, 193-206.
- Levinson, S., Meira, S., & the Language and Cognition Group (2003). 'Natural concepts' in the spatial topological domain – adpositional meanings in crosslinguistic perspective: An exercise in semantic typology. *Language*, 79, 485–516.
- Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106, 1126-1177
- Liszkowski, U., Carpenter, M., & Tomasello, M. (2008). Twelve-month-olds communicate helpfully and appropriately for knowledgeable and ignorant partners. *Cognition*, 108, 732-739.
- McNamara, T. (1986). Mental representations of spatial relations. *Cognitive Psychology*, 18, 87-121.
- Munnich, E., Landau, B., & Doshier, B.A. (2001). Spatial language and spatial representation: a cross-linguistic comparison. *Cognition*, 81, 171-208
- Piantadosi, S., Tily, H., & Gibson, E. (2011). Word lengths are optimized for efficient communication. *PNAS*, 108, 3526-3529.
- Piantadosi, S., Tily, H., & Gibson, E. (2012). The communicative function of ambiguity in language. *Cognition*, 122(3), 280-291.

- Regier, T., Kay, P., & Khetarpal, N. (2007). Color naming reflects optimal partitions of color space. *PNAS*, *104*, 1436-1441.
- Regier, T., Khetarpal, N., & Majid, A. (2013). Inferring semantic maps. *Linguistic Typology*, *17*, (pp. 89-105).
- Shriberg, E., Bates, R., Taylor, P., Stolcke, A., Jurafsky, D., Reis, K., Coccaro, N., Martin, R., Meteer, M., & van Ess-Dykema, C. (1998). Can prosody aid in the automatic classification of dialog acts in conversational speech?. *Language and Speech*, *41*, 439-487.
- Sperber, D., & Wilson, D. (1996). *Relevance: Communication and cognition*. Wiley-Blackwell, 978-0631198789
- Stolcke, A., Reis, K., Coccaro, N., Shriberg, E., Bates, R., Jurafsky, D., Taylor, P., Martin, R., Meteer, M., & van Ess-Dykema, C. (2000). Dialog act modeling for automatic tagging and recognition of conversational speech. *Computational Linguistics*, *26*, 339-371.
- Traxler, M., Pickering, M., & Clifton, C. (1998). Adjunct attachment is not a form of lexical ambiguity resolution. *Journal of Memory and Language*, *39*, 558-592.
- Van Gompel, R., Pickering, M., Pearson, J., & Liversedge, S. (2005). Evidence against competition during syntactic ambiguity resolution. *Journal of Memory and Language*, *52*, 284-307.
- Van Gompel, R., Pickering, M., & Traxler, M. (2001). Reanalysis in sentence processing: Evidence against current constraint-based and two-stage models. *Journal of Memory and Language*, *45*, 225-258.
- Xu, F., & Tenenbaum, J. (2007) Word learning as Bayesian inference. *Psychological Review*, *114*(2), 245-272.