UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

The Pragmatics of Spatial Language

Permalink

Journal

Authors

Ullman, Tomer D.
Xu, Yang
Goodman, Noah D.

Publication Date

2016

Peer reviewed

The Pragmatics of Spatial Language

Tomer D. Ullman (tomeru@mit.edu)
Department of Brain and Cognitive Science,
MIT

Yang Xu (yang_xu_ch@berkeley.edu)
Department of Linguistics, Cognitive Science Program,
UC Berkeley

Noah D. Goodman (ngoodman@stanford.edu)
Department of Psychology, Stanford University

Abstract

How do people understand the pragmatics of spatial language? We propose a rational-speech act model for spatial reasoning, and apply it to the terms 'in' and 'near'. We examine people's fine-grain spatial reasoning in this domain by having them locate where an event occurred, given an utterance. Our pragmatic listener model provides a quantitative and qualitative fit to people's inferences. **Keywords:** Pragmatics, Implicature, Spatial Language

Introduction

Space is continuous, language is discrete, and spatial language is coarse and limited—built from a restricted and closed class of spatial prepositions (Talmy, 1983, 2000; Landau & Jackendoff, 1993) such as "in," "on," and "near." How can we communicate accurately about spatial relations with an impoverished, discrete spatial vocabulary? A partial solution lies in the pragmatics of spatial language. Pragmatic enrichment allows coarse fixed meanings to gain useful context-specific refinements (Grice, 1975; Horn, 1984). This is especially useful when the conveyed states are much finer-grained than the literal vocabulary. Moreover, the spatial domain provides a useful test of pragmatic theory: there is a great deal of room for enrichment in a fine-grained domain.

To illustrate the potential pragmatics of spatial language, consider the top half of Figure 1: this is a map of a small city with two quarters (represented by the red and blue rectangles) and a plaza that is located inside one of the quarters (represented by the dashed circle). Suppose you were told that "a gold lily grew in the red quarter." Where would you think the flower had grown? Taking "in" at face value (i.e. the literal meaning) would mean uniform uncertainty over the red region. But a pragmatic listener could arrive at a more specific interpretation: The speaker did not say the lily was in the plaza, nor near the plaza, nor on the edge of the red quarter.... From this a listener could infer that the lily was probably in none of these locations, and derive a much more specific guess as to where the lily was. In many ways this is a standard scalar implicature (Horn, 1984), such as "some" implying "not all," but the interpretation space is much richer and the effect of context is easily manipulated—if the plaza were not inside the red quarter, or placed differently inside it, a pragmatic listener's interpretation should change.

Because of the fine-grained space of interpretations, spatial language provides a particularly good domain to explore quantitative models of natural language pragmatics, such as the recently successful Rational Speech Acts (RSA)

framework (Frank & Goodman, 2012; N. D. Goodman & Stuhlmüller, 2013). Previous empirical work has discussed the role of pragmatics in spatial locative expressions (Herskovits, 1985, 1987). However, formal work in this domain is scarce and recent computational studies have emphasized production over comprehension (Carstensen, Kon, & Regier, 2014; Golland, Liang, & Klein, 2010). Thus, no studies have looked at the quantitative effects of implicature in the spatial domain. We aim to bridge this gap with a formal RSA-based model, that quantitatively predicts how a listener interprets spatial language in a 2D map domain.
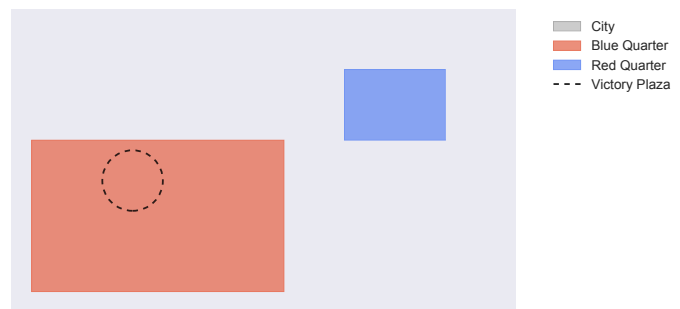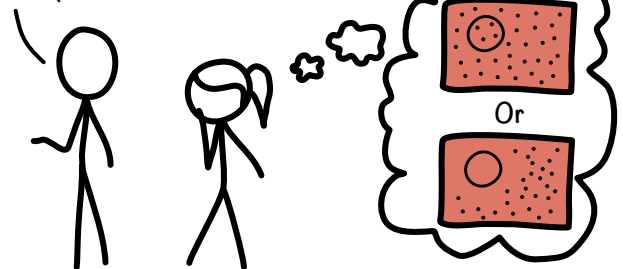

Figure 1: Illustration of a simple spatial reasoning situation. How will the listener interpret the speaker's words?

Modeling spatial implicature

We model spatial language understanding within the Rational Speech Acts (RSA) framework of Frank and Goodman (2012). This framework has previously been applied to domains such as object properties, with terms such as quantifiers (e.g. "some of the apples are red" implying that only a subset of the apples are red). We adapt this framework to spatial language by taking locations as the domain of interest, and

building a lexicon for describing spatial relations.

The RSA framework specifies a pragmatic listener reasoning about the intention of an informative speaker, who in turn reasons about a literal listener; the literal listener updates her beliefs by conditioning on the truth of an utterance. Altogether, writing the location of interest (e.g. of the gold lily) as s and the utterance (e.g."in the red quarter") as u:

$$P_{L_0}(s|u) \quad \propto \quad \delta_{[\![u]\!](s)} P(s) \tag{1}$$

$$P_{S_1}(u|s) \quad \propto \quad \sum_{s' \text{ s.t. } |s-s'|<\epsilon} P_{L_0}(s'|u) P(u) \tag{2}$$

$$P_{L_1}(s|u) \quad \propto \quad P_{S_1}(u|s) P(s) \tag{3}$$

Beginning with the simplest listener model, L_0, Equation 1 specifies a listener who interprets the utterance via its truth-functional denotation $[\![u]\!]$ (described in more detail below), and simply restricts the prior distribution over locations to those where this denotation is true. The speaker, S_1, in Equation 2 aims for the literal listener to guess the correct location, balanced against the prior probability of the utterance. However, because it is vanishingly unlikely for L_0 to guess exactly the right location, our speaker only cares if the true and gusseded locations are within a small distance ϵ; this corresponds to an *approximate question under discussion* as used by Kao, Wu, Bergen, and Goodman (2014). Finally, the pragmatic listener, L_1, in Equation 3 updates her beliefs in accord with Bayes' rule, under the assumption that the S_1 speaker would have chosen the utterance she heard.

The denotation of an utterance u is a function from location to Boolean: $[\![u]\!](s) \in$ Bool. We will assume that the spatial language relevant to our scenarios depends on a context set of regions in $\mathbb{R}^2$, as in Figure 1: The City, The Red Quarter, The Blue Quarter, The Victory Plaza.[1] We considered three types of spatial utterances, each of which combines a preposition with a region name:

- "In" utterances (e.g. "In the Red Quarter") are true within the region: $[\![\text{"in R"}]\!](s) := s \in R$.

- "Near edge" utterances (e.g. "Near the edge of the Red Quarter") are true within distance d of a region boundary: $[\![\text{"near edge R"}]\!](s) := \mu(s, \beta(R)) < d$. (Where μ measures distance to a set, and β returns the boundary of a region.)

- "Near" utterances (e.g. "Near the Red Quarter") are true near the edge but not inside: $[\![\text{"near R"}]\!](s) := [\![\text{"near edge R"}]\!](s) \wedge \neg [\![\text{"in R"}]\!](s)$.

People no doubt have access to many more spatial utterances, and their combinations. We restrict ourselves to a small set of utterances for simplicity, and to these specific utterances because they were studied in previous research (Herskovits, 1985), and because a separate production task (not reported here) shows these are common spatial terms. We set a uniform prior over the utterances, $P(u)$, and location, $P(s)$.

The model as stated leaves open the question of what distance counts as "Near" or "Near edge" in the shared lexicon (what d to use), as well as what counts as "approximately equal" between the true state and the inferred state (what ϵ to use). We set ϵ to 20 units (pixels) based on initial exploration. Because the flexibility of "Near" seems to depend on the item and context ("near your coffee mug" is not the same as "near your coffee house", see also Burigo & Coventry, 2010), we allow the literal listener to establish the most useful d for each utterance. We extend Equation 1 to include uncertainty about the tolerances for each "near" utterance:

$$P_{L_0}(s|u) \propto \sum_{\vec{d}} \delta_{[\![u]\!]^{\vec{d}}(s)} P(s) P(\vec{d}), \tag{4}$$

where $\vec{d}$ is a real number, the tolerance. We assume $P(\vec{d})$ is uniform over 10-30 units (pixels) for each utterance[2].

We implement this model in the probabilistic programming language Church (N. Goodman, Mansinghka, Roy, Bonawitz, & Tarlow, 2012). A full implementation can be found at `http://forestdb.org/models/spatialImplicature.html`. We used this implementation to generate model predictions for each spatial configuration ("city map") and utterance used in the experiment below.

Experiment

We examined people's spatial inferences and the predictions of our model for the spatial terms 'in' and 'near', by putting participants in the role of a listener and asking them to guess where an event happened on a map. Participants made their guess in response to the utterance of a speaker with access to the location of the event.

Participants, materials and methods

Participants ($N = 49$, 13 female, median age 29) were recruited through Amazon's Mechanical Turk service.

We created 4 city maps like the one shown in Figure 1, each containing 2 "Quarters" of different size and color, and a circle labeled "Victory Plaza". The location of the Plaza varied between the cities, while the location of the Quarters remained the same. To broadly control for color and position, we created another set of 4 maps by flipping the original maps along the vertical and horizontal axis, and changing the color of the quarters, making 8 maps in total. Participants were randomly assigned to one of the two map groups.

Participants were told that they would see a series of city maps, that a special flower called a 'Gold Lily' can grow anywhere in the city they're shown, and that their task is to find the Gold Lilies.

Participants were further told that for each map a person will tell them where a Gold Lily grew, and that this person can say the Gold Lily grew *in* a location or *near* a location (this narrows people's lexicon to one more directly comparable with the model). This person was said to be reasonable,

---

[1]Both the denotation and Equations 1–3 depend on the context of named regions. We leave this dependence implicit to avoid clutter.

[2]For comparison, the maps used are 500x300 units (pixels).

IN…Red/Blue/Plaza/City
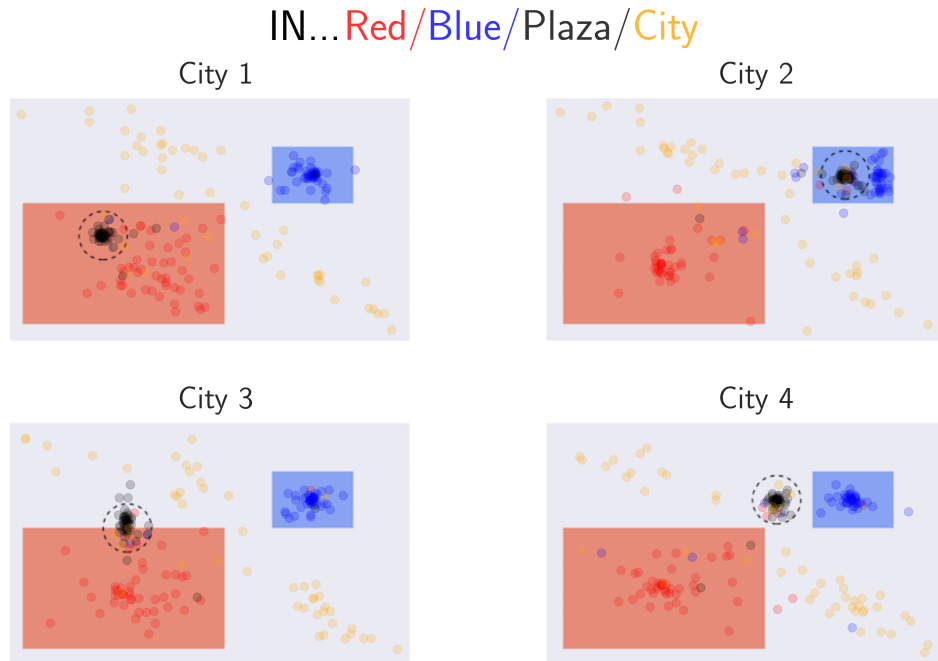
City 1

City 2

City 3

City 4

Figure 2: Experiment results pooled across participants for 'In' utterances. Each dot is a participant guess for where a Gold Lily grew, color coded by the region that 'In' refers to (e.g. red dots are different responses to "In the Red Square").

NEAR…Red/Blue/Plaza

City 1

City 2

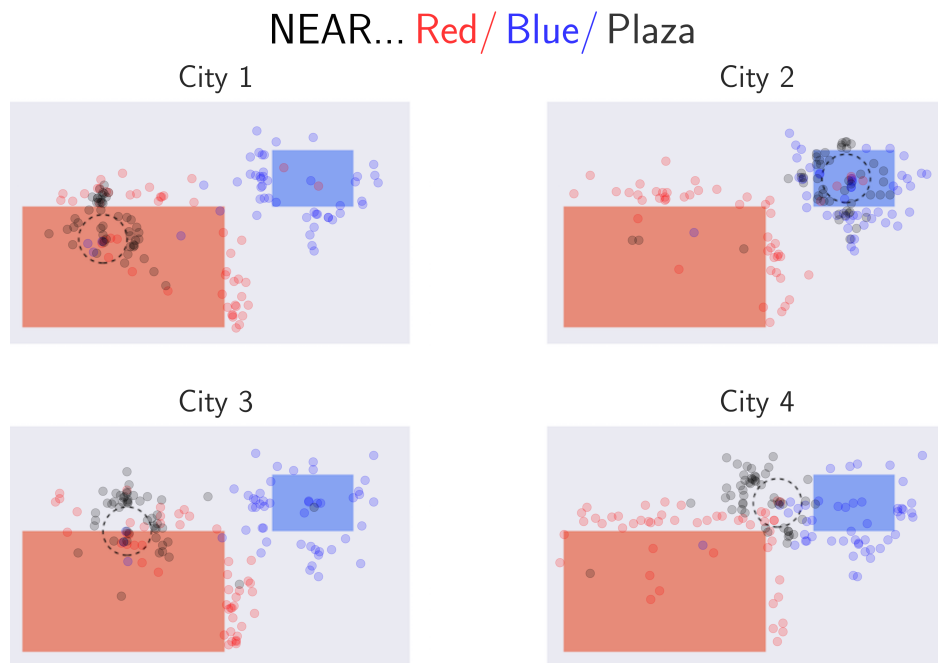

City 3

City 4

Figure 3: Experiment results pooled across participants for 'Near' utterances, details as in Fig. 2.

and honest. As an example, a participant might read the sentence 'A person tells you: A Gold Lily grew in the Red Quarter'. Participants made their guess by clicking directly on the maps they were shown.

For each of the 4 maps in their group, participants were prompted with a sentence made of *word × location*

combinations, where $word \in [In, Near]$ and $location \in [Red Quarter, Blue Quarter, Victory Plaza, City]$. The combination "Near the City" was not used, as this area was not in the scope of the image and likely to create confusion. In total, each participant was prompted with 28 sentences in random order. Each map had a legend to its right.
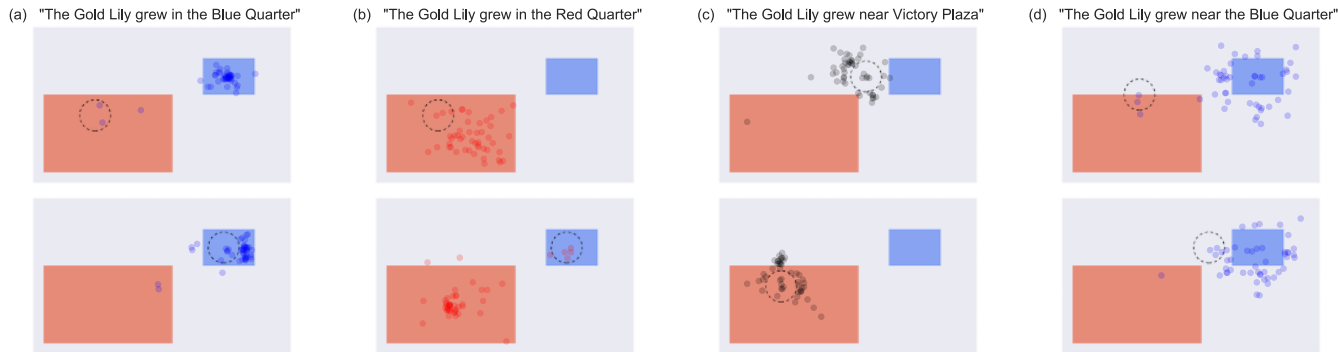
1615

Figure 4: Examples highlighting trends from Figures 2 and 3. The response to the same utterance (column) changes by the different relative configuration of the regions.

Qualitative Results

Figures 2 and 3 show responses for "in" and "near"sentences, collapsing across the rotated-and-inverted-color cities and color-coded by the region referred to. For example, in the top-left of Figure 2, the blue dots correspond to all participant guesses for where the flower grew in City 1, when prompted with "In the Blue Quarter".

To highlight qualitative effects, we focus on the cases highlighted in Figure 4. The reader is encouraged to examine the remaining twenty cases, which show similar effects.

'In X' implies 'In X but not in Y' As shown for example in Figure 4(a): Both the top and bottom of (a) show guesses for lily location when told it grew in the Blue Quarter. When people hear 'In Blue' in City 1 (top) they infer 'In center of Blue'. But in City 2 (bottom) they infer 'In Blue, but not in the Plaza'. Specifically, in City 1 (top) a tic-tac-toe-like division of the Quarter shows that the grid center, accounting for 11% of the area, captures 59% of the guesses. In City 2 (bottom) the same grid center contains just 8% of the guesses, as people shift guesses to the right of the Plaza. Such a pattern of results holds for the other regions as well (Figure 2).

Edge avoidance When there is no direct intersection between regions (except the City, which all regions intersect with), people do not guess uniformly in the region, but rather favor the center. For example, in Figure 4(b), bottom: People told 'In Red' place most of their guesses in the center of the region. A tic-tac-toe grid-analysis shows the red-grid center accounts for 11% of the area but 65% of the guesses. In 10,000 simulations of 49 participants that guess uniformly over the Red Quarter, *none* have the center account for more than ∼30% of the responses. This pattern of results holds for the other regions as well, as shown in Figure 2, though it is disrupted when the Plaza is placed inside the region as in the top panel of Figure 4(b).

'Near' is non-uniform on edges As shown in Figure 4(c) and (d), when told 'Near Plaza' or 'Near Blue' people do not seem to be guessing uniformly, rather their guesses depend on the context of other regions. In the top panel of (c) people place their guesses for 'Near Plaza' to the top-left of

the Plaza, while in the bottom panel of (c) people place their guesses for 'Near Plaza' to the top and right of the plaza. In the top panel of (d) the guesses for 'Near Blue' are to the left and bottom of the Blue Quarter, while in bottom of (d) they shift away from the left of the Blue Quarter (now partly occupied by the Plaza). Figure 5 shows the percentage of participant guesses by quadrant, indicating the by region under consideration. A χ^2 test on binned responses near each region (as in Figure 5) shows 11 of the 12 response-patterns are significantly different from uniform ($p < 0.05$), with the Blue Quarter in City 3 being the only exception.

This qualitative pattern of results make it clear that context affects participants' interpretations of spatial language, as expected under a pragmatic account. In the next section, we turn to a quantitative analysis, comparing results with the different listener models described above.

Comparison to Model(s)

To quantitatively compare people to the literal and pragmatic listener models (L_0 and L_1), we converted participants' responses for each condition into smooth two-dimensional dis-
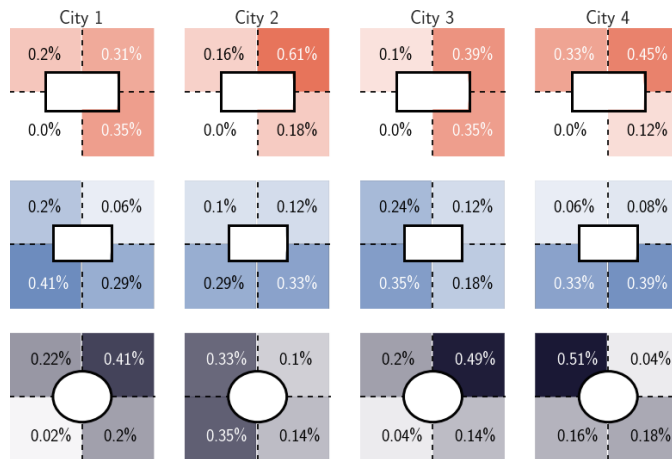

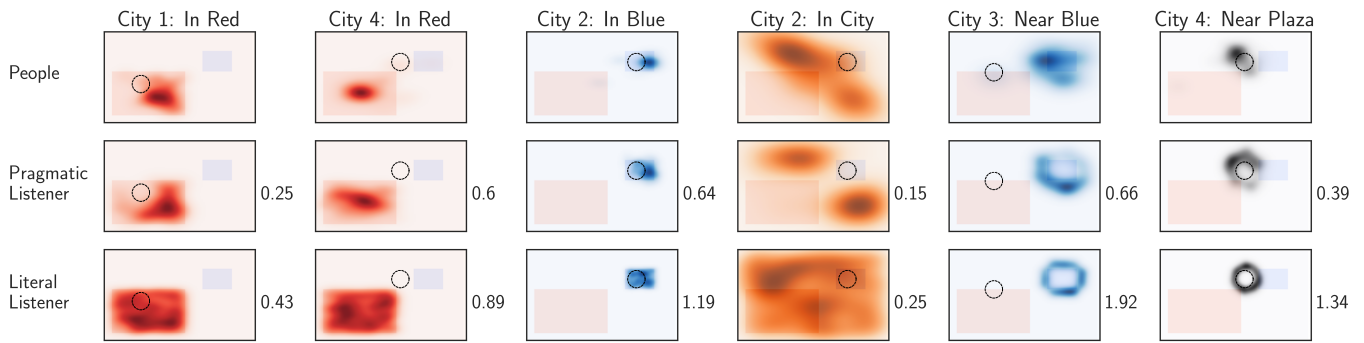Figure 5: Distribution of participant 'Near' guesses, by region (color-coded) and city.

Figure 6: Example comparisons between people and the two listener models. Colored patches indicate probability distributions inferred from people's responses and model samples. Numbers to the right of each listener subplot indicate the *KL* distance between the distributions of people and the model samples, lower numbers indicate a better match with people.

tributions. We used non-parametric, multivariate kernel density estimation to infer these distributions (Epanechnikov, 1969). The same estimation was applied to samples drawn from the probabilistic programs that represent the L_0 and L_1 models. The model distributions are 'sharpened' by raising them to a power α and renormalizing. Such a sharpening parameter controls how much of the model's distribution is smeared or centered around the maximum likelihood, and is functionally similar to an 'optimality' parameter used in other implicature studies (N. D. Goodman & Stuhlmüller, 2013; Kao et al., 2014). We set $\alpha = 1.5$, although the overall findings are robust within an explored range $\alpha \in [1, 2]$.

Example comparisons between the models and people are shown in Figure 6. When hearing 'In Red' in City 1 (left-most column) The literal listener L_0 places a uniform posterior distribution on the Red Quarter, while L_1 and people avoid Victory Plaza and lean to the right of the Red Quarter. This correspondence can be measured in terms of the Kullback-Leibler (*KL*) distance between the distributions. In this particular example, $KL(\text{people}|L_0) = 0.43$, while $KL(\text{people}|L_1) = 0.25$.

As Figure 6 shows, the L_1 model captures many of the interesting spatial-implicature interpretations shown by people. Indeed, the L_1 listener shows the same general qualitative patterns discussed in the previous section ("in X" implies "in X but not Y", edge avoidance, non-uniformity on edges). A full comparison figure exceeds the space limits of this paper, but can be found at http://www.mit.edu/~tomeru/spatialPragmatics/allComparisons.pdf. Figure 7 shows the *KL* distances for both models across all questions and cities. The pragmatic listener is closer to people than the literal listener for 26 of the 28 (93%) comparisons.

While the pragmatic listener is able to account for the pattern of results found with people, it is interesting to highlight some particular cases where it does not do so well. The left column of Figure 8 shows one of the two cases where the *KL* distance between people and the literal listener is shorter than between people and the pragmatic listener. While the implicature pattern of the pragmatic listener seems generally correct – it places its probability in the Red Quarter while avoiding the Plaza – the model is over-avoiding the area near the plaza relative to people. The right column of

Figure 8 shows an example where even though the pragmatic listener is closer to people than the literal listener, it seems generally mistaken in that it places much of its probability distribution within the Red Quarter, which people avoid.

Overall, the results suggest that the pragmatic listener model can account for the quantitative and qualitative pattern of pragmatic spatial inferences of people, within our domain.

Discussion

Space is fine grained, more so than language, leaving a great deal of room for the influence of context and pragmatic enrichment. As a result, spatial language is a promising avenue for testing quantitative models of pragmatic reasoning. We investigated people's fine-grained pragmatic reasoning in a spatial domain, and found support for a model within the rational speech act (RSA) framework.

The model's lexicon was more restricted than natural spatial language. Even within our domain, utterances might include 'in the middle of the City', 'in between the Red and Blue Quarters', 'above Victory Plaza', 'to the left of the Blue Quarter, near the Plaza'.... In future work, a parallel production task (in which people convey a map location using language) should be used to inform and expand the lexicon.

There are many other extensions of the paradigm. For instance, when reasoning pragmatically, people take into account the knowledge of their partner (N. D. Goodman & Stuhlmüller, 2013). In a spatial context, a speaker might adjust the landmarks and utterances they use depending on the listener: 'next to Jacques' café' for those in the know vs. 'next to the Louvre' for out-of-towners. The listener could correspondingly adjust their inferences depending on what they believe the speaker knows.

Within the model, an important directions for exploration is the treatment of *pragmatic slack* in phrases like "near the Plaza." We modeled the tolerance of the lexical items as uncertain, letting the literal listener infer an appropriate value. This literal-slack model allows the slack to adjust in order to increase the odds of a speaker's utterance being true; it neglects other potentially important pressures, such as making the speaker's utterance *informative*. One option is *lifting* the
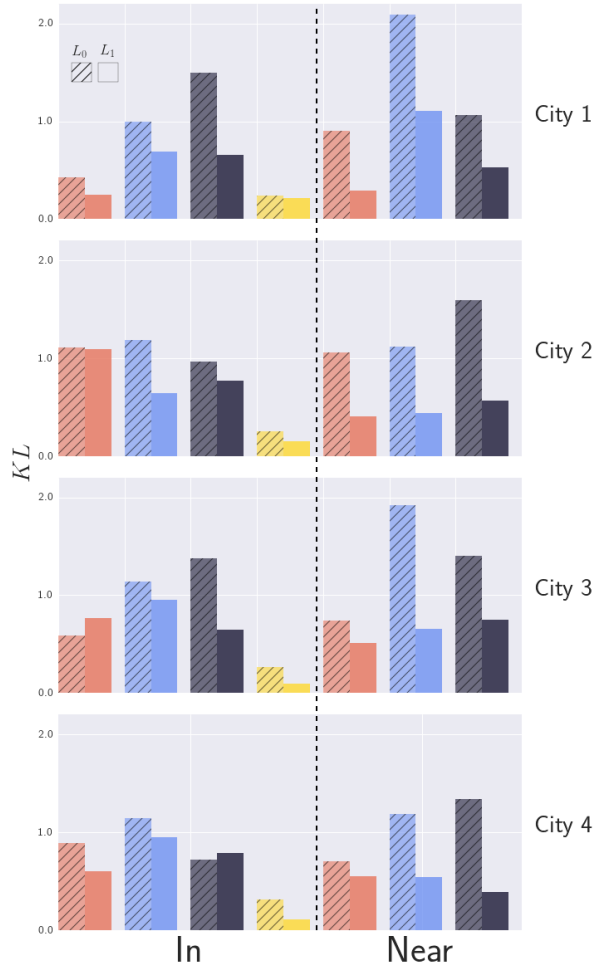
Figure 7: KL distances between people and the listener models, by question and city. Hatched bars are literal listener L_0 and plain bars are pragmatic listener L_1. Bar color indicates region. Lower bars indicate greater agreement with people.
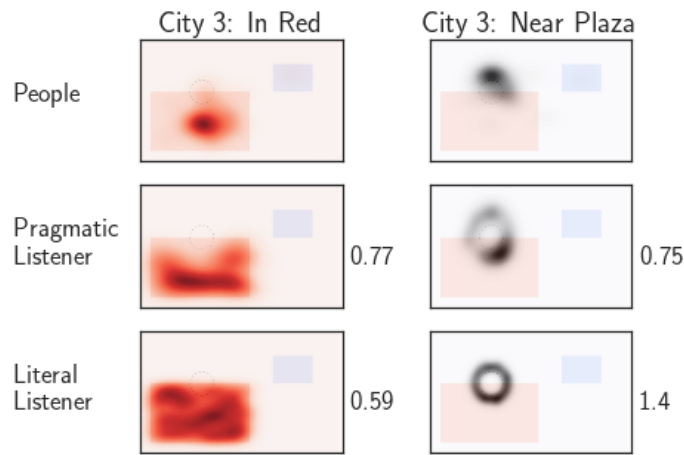

Figure 8: Mismatches between pragmatic listener and people, with literal listener for comparison. Details similar to Fig. 6.

slack variable into the pragmatic listener (as in N. D. Goodman & Lassiter, 2014). While this model is simple to construct, future work is needed to find cases which distinguish it from the model used here, possibly in a communicative paradigm with participant as both speaker and listener.

References

Burigo, M., & Coventry, K. (2010). Context affects scale selection for proximity terms. *Spatial Cognition & Computation*, *10*(4), 292–312.

Carstensen, A., Kon, E., & Regier, T. (2014). Testing a rational account of pragmatic reasoning: The case of spatial language. , 2009–2013.

Epanechnikov, V. A. (1969). Non-parametric estimation of a multivariate probability density. *Theory of Probability & Its Applications*, *14*(1), 153–158.

Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, *336*(6084), 998–998.

Golland, D., Liang, P., & Klein, D. (2010). A game-theoretic approach to generating spatial descriptions. In *Proceedings of the 2010 conference on empirical methods in natural language processing* (pp. 410–419).

Goodman, N., Mansinghka, V., Roy, D., Bonawitz, K., & Tarlow, D. (2012). Church: a language for generative models. *arXiv preprint arXiv:1206.3255*.

Goodman, N. D., & Lassiter, D. (2014). Probabilistic semantics and pragmatics: Uncertainty in language and thought. *Handbook of Contemporary Semantic Theory. Wiley-Blackwell*.

Goodman, N. D., & Stuhlmüller, A. (2013). Knowledge and implicature: Modeling language understanding as social cognition.

Grice, P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics, vol.3* (pp. 41–58). Academic Press.

Herskovits, A. (1985). Semantics and pragmatics of locative expressions*. *Cognitive Science*, *9*(3), 341–378.

Herskovits, A. (1987). *Language and spatial cognition*. Cambridge University Press.

Horn, L. (1984). Toward a new taxonomy for pragmatic inference: Q-based and r-based implicature. *Meaning, form, and use in context: Linguistic applications*, 11–42.

Kao, J. T., Wu, J. Y., Bergen, L., & Goodman, N. D. (2014). Nonliteral understanding of number words. *Proceedings of the National Academy of Sciences*, *111*(33), 12002–12007.

Landau, B., & Jackendoff, R. (1993). Whence and whither in spatial language and spatial cognition? *Behavioral and Brain Sciences*, *16*(02), 255–265.

Talmy, L. (1983). *How language structures space*. Springer.

Talmy, L. (2000). *Towards a cognitive semantics, Volume 1: Concept Structuring Systems.* MIT Press.