

The Pragmatics of Metaphor Understanding: A Computational Approach

Justine T. Kao¹ (justinek@stanford.edu), Leon Bergen² (bergen@mit.edu), Noah D. Goodman¹ (ngoodman@stanford.edu)

¹Department of Psychology, Stanford University. ²Department of Brain and Cognitive Science, MIT.

Abstract

Metaphor understanding has been widely studied in psychology, linguistics, philosophy, and computer science. One account of metaphor understanding is pragmatics. Some metaphor interpretations arise from pragmatics. Here we present a computational model of pragmatics that accounts for blah blah blah Metaphor understanding has been widely studied in psychology, linguistics, philosophy, and computer science. One account of metaphor understanding is pragmatics. Some metaphor interpretations arise from pragmatics. Here we present a computational model of pragmatics that accounts for blah blah blah

Keywords: language understanding; metaphor; pragmatics; computational models

Introduction

From “Juliet is the sun” to “No man is an island,” nonliteral language is, quite literally, everywhere. Human communication is laden with metaphor and hyperbole, often creating poetic or humorous effects that add rich and important dimensions to language (Glucksberg, 2001; Pilkington, 2000; Lakoff & Turner, 2009; Roberts & Kreuz, 1994; B. Bergen & Binsted, 2003). Metaphor has inspired a particularly abundant amount of research in cognitive science, spanning topics such as how metaphors structure and shape our thoughts (Ortony, 1993; Lakoff et al., 1993; Thibodeau & Boroditsky, 2011), whether metaphor processing recruits the same strategies as standard language processing (Giora, 1997; Ortony, Schallert, Reynolds, & Antos, 1978; Gibbs Jr, 2002; Glucksberg & Keysar, 1993) and what factors determine the meaning and aptness of a novel metaphor (Blasko & Connine, 1993; Tourangeau & Sternberg, 1981; Kintsch & Bowles, 2002). The overwhelming interest in metaphor research is due in part to the ubiquity of metaphor in everyday language as well as the belief that metaphor may be critical for helping us understand how the mind creates meaning.

One approach to studying metaphor focuses on the pragmatic principals that listeners utilize to infer meaning from metaphorical utterances (Tendahl & Gibbs Jr, 2008; Stern, 2000). Rather than view metaphor as a separate mode of communication that requires specialized language processing strategies, this approach argues that basic principles of communication drive the meaning that a listener infers from a metaphor (Sperber & Wilson, 2008). Relevance theory, in particular, posits that listeners interpret utterances with the assumption that speakers produced them because they are maximally relevant. Relevance theorists argue that this principle explains how listeners infer the meaning of a novel metaphor

as well as other forms of loose talk where the meaning of the utterance is underspecified (Wilson & Sperber, 2002; Wilson & Carston, 2006; Sperber & Wilson, 1985). When interpreting the metaphor “My lawyer is a shark,” for example, the listener assumes that the speaker aims to communicate features of “a shark” that are relevant to the person under discussion (“my lawyer”), and thus do not access shark features such as *has fins* or *swims*.

While many linguists and psychologists have argued for the benefits of studying metaphor using a pragmatics framework, to our knowledge there is no formal model showing that effects in metaphor understanding may arise from basic principles of communication. On the other hand, a recent body of work presents a series of computational models for pragmatic reasoning, where speaker and listener reason about each other to communicate effectively (Frank & Goodman, 2012; Jäger & Ebert, 2009). By formalizing principals of communication, these Rational Speech Act models are able to quantitatively explain a range of phenomena in language understanding, such as scalar implicature and the effect of alternative utterances (Goodman & Stuhlmüller, 2013; L. Bergen, Goodman, & Levy, 2012). However, a limitation of these models is that they are unable to predict interpretations of an utterance that are false under its literal meaning. In this paper, we extend the model to consider communicative goals established by context that may be optimally satisfied by metaphorical utterances. A listener assumes that the speaker chooses an utterance to maximize informativeness about a subject along dimensions that are relevant to the conversation. This makes it possible for a literally false utterance to be optimal as long as it is informative along the target dimension. This framework closely aligns with the relevance-theoretic view that a listener considers the relevance of a meaning to the question under discussion in order to infer what the speaker intended to communicate.

Although metaphor understanding is a complex phenomenon that calls for a variety of approaches, we present a computational model to argue that the interpretation of at least some types of metaphor are shaped at least in part by basic principals of pragmatics. To reasonably limit the scope of our work, we focus on metaphors of the classic form “*X* is a *Y*.” We describe a computational model that can interpret such sentences metaphorically and conduct behavioral experiments to evaluate the model’s performance. We show that a listener’s interpretation of a metaphor is driven by context and the question under discussion, and that this effect is captured by our formalization of maximizing informative with respect to the communicative goal. [Finally, we show that metaphors often communicate more information than literal statements and hence can be optimal and rational speech acts. Or what-

ever we end up focusing on in the error analysis]

Computational Model

At the core of basic Rational Speech Act models, a listener and a speaker recursively reason about each other to arrive at pragmatically enriched meanings. Given an intended meaning, a speaker reasons about a literal listener and chooses an utterance based on its informativeness. A pragmatic listener then reasons about the speaker and uses Bayes’s Rule to infer the meaning given the utterance. To account for nonliteral interpretation, we extend this model by considering the idea that a speaker may have a range of different communicative goals (). Intuitively, an utterance is optimally informative and relevant if it satisfies the speaker’s communicative goal. Since the speaker’s precise communicative goal may be unknown to the listener, the listener performs joint inference on the goal as well as the intended meaning. By introducing multiple potential goals for communication, we open up the possibility for a speaker to produce an utterance that is literally false but still satisfies her goal. The speaker achieves this in part by exploiting her and the listener’s prior knowledge—their common ground (Clark, 1996)—to reason about what information the listener would gain if he takes the utterance literally.

To illustrate this idea and demonstrate how it is implemented in our model, we will use the metaphor “John is a shark” as an example. For simplicity, in this model we restrict the number of possible categories to which a member may belong to c_a and c_p , denoting an animal category or a person category, respectively. We also restrict the possible features of John under consideration to a vector of size three: $\vec{f} = [f_1, f_2, f_3]$, where f_i is either 0 or 1. We denote a possible communicative goal that a speaker may have as $g_{i,j}$, where $1 \leq i \leq 3$ and $j \in \{0, 1\}$. $g_{i,j}$ is a function such that $g_{i,j}(\vec{f}) = 1$ if $f_i = j$, and 0 otherwise.

Suppose the speaker S_1 ’s goal is to communicate that John has feature $f_1 = 1$, where f_1 is the property *scary*. S_1 reasons about a literal listener L_0 who will interpret an utterance “John is a shark” as meaning that John is literally a member of the category “shark.” The literal listener is modeled as:

$$L_0(c, \vec{f}|u) = \begin{cases} P(\vec{f}|c) & \text{if } c = u \\ 0 & \text{otherwise} \end{cases}$$

where $P(\vec{f}|c)$ is the prior probability that a member of a category c (in this case “shark” or “person”) has feature values \vec{f} .

Given her goal, S_1 chooses an utterance u according to a softmax decision rule that describes an approximately rational planner, where $U(u|g)$ is the speaker’s utility of producing u given she has goal g (Sutton & Barto, 1998).

$$S_1(u|g) \propto e^{\lambda U(u|g)} \quad (1)$$

The utility U of an utterance given a goal is composed of both the probability of the speaker’s goal being satisfied

and the negative of the utterance cost. Optimizing the probability of the speaker’s goal being satisfied can be accomplished by minimizing the goal’s information-theoretic surprisal given an utterance. Since given an utterance u the listener L_0 will guess that the meaning is c, \vec{f} with probability $L_0(c, \vec{f}|u)$, the probability of the speaker’s goal being satisfied is $\sum_{c, \vec{f}} L_0(c, \vec{f}|u) g(\vec{f})$. Using the information-theoretic definition of surprisal and a uniform utterance cost, equation (1) becomes the following, where λ is an optimality parameter that we fit to the data:

$$S_1(u|g) \propto \lambda \sum_{c, \vec{f}} L_0(c, \vec{f}|u) g(\vec{f}) \quad (2)$$

Intuitively, based on S_1 ’s understanding of L_0 ’s prior knowledge, she knows that if she produces the utterance “John is a shark,” L_0 will believe that John is literally a shark and hence very likely to be *scary*. Since S_1 ’s goal is satisfied if the listener believes that John is scary, S_1 is motivated to produce such a metaphorical utterance.

A pragmatic listener L_1 now reasons about such a speaker. Based on prior knowledge, L_1 knows that John is extremely unlikely to be literally a member of the shark category. On the other hand, L_1 knows that the speaker S_1 is fairly likely to want to communicate about John’s scariness. L_1 also knows that S_1 knows that *scary* is a high-probability feature of sharks. The listener L_1 performs Bayesian inference to guess the intended meaning given prior knowledge and his internal model of the speaker. To determine the speaker’s intended meaning, L_1 will marginalize over the possible speaker goals under consideration.

$$L_1(c, \vec{f}|u) \propto \sum_g P(c) P(\vec{f}|c) P(g|\vec{f}) S_1(u|g)$$

Given L_1 ’s prior knowledge, his model of the speaker, and the utterance she produces, he infers that the meaning of the utterance is likely to be that John is a scary person. Note that while speaker and listener can continue to recursively reason about each other indefinitely, in this paper we present interpretation results for L_1 . To arrive at an interpretation, the listener needs to consider the following prior probabilities:

- (1) $P(c)$, the prior probability that X belongs to category c . We assume that L_1 is extremely certain that the person under discussion (e.g. John) is a person, but that there is a non-zero probability that John is actually a non-human animal. We fit $P(c_a)$ to data with the assumption that $10^{-4} \leq P(c_a) \leq 10^{-1}$.
- (2) $P(\vec{f}|c)$, the prior probability that a member of category c has feature values \vec{f} . We obtain this empirically in Experiment 1.
- (3) $P(g|\vec{f})$, the prior probability that given that a speaker knows the value of the feature vector \vec{f} , she wishes to communicate goal g . This prior can change given the context that a question sets up. For example, if S_1 is responding to the question “Is John scary?,” she is much more

Table 1: This is a table

Animal	f1=1	f1=0	f2=1	f2=0	f3=1	f3=0	Animal	f1=1	f1=0	f2=1	f2=0	f3=1	f3=0
ant	small	large	strong	weak	busy	idle	goose	loud	quiet	mean	nice	annoying	agreeable
bat	scary	unalarmin	blind	sighted	nocturnal	diurnal	horse	fast	slow	strong	weak	beautiful	ugly
bear	scary	unalarmin	big	small	fierce	nonviolent	kangaroo	jumpy	relaxed	bouncy	inelastic	cute	unattractive
bee	busy	idle	small	large	angry	unangry	lion	ferocious	nonviolent	scary	unalarmin	strong	weak
bird	free	unfree	graceful	awkward	small	large	monkey	funny	humorless	smart	stupid	playful	unplayful
buffalo	big	small	strong	weak	wild	tame	owl	wise	foolish	quiet	loud	nocturnal	diurnal
cat	independent	dependent	lazy	fast	soft	hard	ox	strong	weak	big	small	slow	fast
cow	fat	thin	dumb	smart	lazy	fast	penguin	cold	hot	cute	unattractive	funny	humorless
dog	loyal	disloyal	friendly	unfriendly	happy	unhappy	pig	dirty	clean	fat	thin	smelly	fragrant
dolphin	smart	stupid	friendly	unfriendly	playful	unplayful	rabbit	fast	slow	furry	hairless	cute	unattractive
duck	loud	quiet	cute	unattractive	quacking	non-quacking	shark	scary	unalarmin	dangerous	safe	mean	nice
elephant	huge	small	smart	stupid	heavy	light	sheep	wooly	hairless	fluffy	hard	dumb	smart
fish	scaly	smooth	wet	dry	smelly	fragrant	tiger	striped	unpatterned	fierce	nonviolent	scary	unalarmin
fox	sly	artless	smart	stupid	pretty	ugly	whale	large	small	graceful	awkward	majestic	inferior
frog	slimy	nonslippery	noisy	quiet	jumpy	relaxed	wolf	scary	unalarmin	mean	nice	angry	unangry
goat	funny	humorless	hungry	full	loud	quiet	zebra	striped	unpatterned	exotic	native	fast	slow

likely to have the goal of communicating John’s scariness. If the question is vague, e.g. “What is John like?”, the prior over goals is uniform. If the question is specific, the prior probability for S_1 ’s goal being to answer that specific question is a parameter p_g that we assume to be greater than 0.5 and fit to data.

Behavioral Experiments

To evaluate our model’s interpretation of metaphor, we focused on a set of 32 metaphors comparing human males to different non-human animals. Experiment 1A and 1B were conducted to elicit prior feature probabilities for the domains in question. We then conducted Experiment 2 to measure people’s interpretations for the set of metaphors.

Experiment 1A: Feature Elicitation

Materials We selected 32 common non-human animal categories from an online resource for learning English (url).

Methods 100 native English speakers with IP addresses in the United States were recruited on Amazon’s Mechanical Turk. Each subject read 32 animal category names presented in random order, e.g. “whale”, “ant”, “sheep”. For each animal category, subjects were asked to type the first adjective that came to mind in a text box.

Results Using subjects’ responses, we constructed a list of adjectives for each animal category and ordered them by the number of times they were given by a different subject (i.e. their popularity). We removed all color adjectives, such as “brown” and “black.” To avoid constructing a set of features that have roughly equivalent meanings such as “big”, “huge”, and “large”, we used Wordnet (Miller, 1995) to identify synonymous adjectives and only kept the most popular adjective among a set of synonyms. We then took the top three most popular adjectives for each animal category and used them as the set of features. In what follows, f_1 is the most popular adjective, f_2 the second, and f_3 the third. Table 1 shows the animal categories and their respective features.

Experiment 1B: Feature Prior Elicitation

Materials Given the features collected from Experiment 1A, we elicit the prior probability of a feature vector given an animal or person category $P(\vec{f}|c)$. We make the simplifying assumption that given a feature f_i under consideration (e.g. *scary*), its antonym (e.g. *unalarmin*) represents the absence of that feature. We used Wordnet to construct antonyms for each of the adjective features produced in Experiment 1A. When multiple antonyms existed or when no antonym could be found on Wordnet, the first author used her judgment to choose the appropriate antonym. Table 1 shows the resulting list of antonyms.

For each animal category, eight possible feature combinations were constructed from the three features and their antonyms. For example, the possible feature combinations for a member of the category “ant” are {small, strong, busy}, {small, strong, idle}, {small, weak, busy}, and so on.

Methods 60 native English speakers with IP addresses in the United States were recruited on Amazon’s Mechanical Turk. Each subject completed 16 trials in random order. Each trial consisted of the eight feature combinations for a particular animal category. Using slider bars with ends marked by “Impossible” and “Absolutely certain,” subjects were asked to rate how likely it is for a member of the animal category to have each of the eight feature combinations. Subjects also rated the probabilities of the feature combinations for a male person. We only elicited priors for males in order to be consistent with Experiment 2 and minimize confounds with gender.

Results We normalized each subject’s ratings for the eight feature combinations in a trial to sum up to 1 based on the assumption that the eight feature combinations exhaustively describe a member of a particular category, . Averaging across subjects’ normalized ratings, we obtained the feature priors $P(\vec{f}|c)$ for $c = c_a$ (animal) and $c = c_p$ (person), where $f_i = 1$ is represented by the feature adjective and $f_i = 0$ is represented by its antonym. Figure 1 shows the average marginal

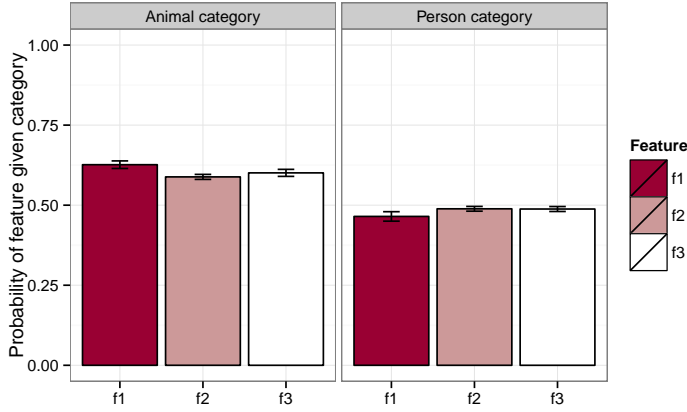


Figure 1: This is a figure.

probabilities of features being present given an animal category versus a person category. Since the features were created using the animal categories in Experiment 1A, by construction features are rated as significantly more likely to be present given the animal category than given the person category. This suggests that subjects are confident that each animal category has certain distinguishing features, while those same features are rated as appearing in people with roughly 0.5 probability on average.

Experiment 2: Metaphor Understanding

Materials We created 32 scenarios based on the animal categories and results from Experiment 1. In each scenario, a person (e.g. Bob) is having a conversation with his friend about a person that he recently met. Since we are interested in how the communicative goals set up by context affect metaphor interpretation as well as the effectiveness of metaphorical versus literal utterances, we created four conditions for each scenario by crossing vague/specific goals and literal/metaphorical utterances. In vague goal conditions, Bob’s friend asks a vague question about the person Bob recently met: “What is he like?” In specific goal conditions, Bob’s friend asks a specific question about the person: “Is he $f1$?” Where $f1$ is the most popular adjective for a given animal category c_a (see Table 1). In literal conditions, Bob replies with a literal utterance, either by saying “He is $f1$ ” to the question “What is he like?” or “Yes” to the question “Is he $f1$?”. In Metaphorical conditions, Bob replies with a metaphorical statement, e.g. “He is a c_a ” where c_a is an animal category. See Table 2 for examples of this experiment scheme.

Methods 49 native English speakers with IP addresses in the United States were recruited on Amazon’s Mechanical Turk. Each subject completed 32 trials in random order. The 32 trials were randomly and evenly assigned to one of the four conditions, i.e. each subject read 8 scenarios for each

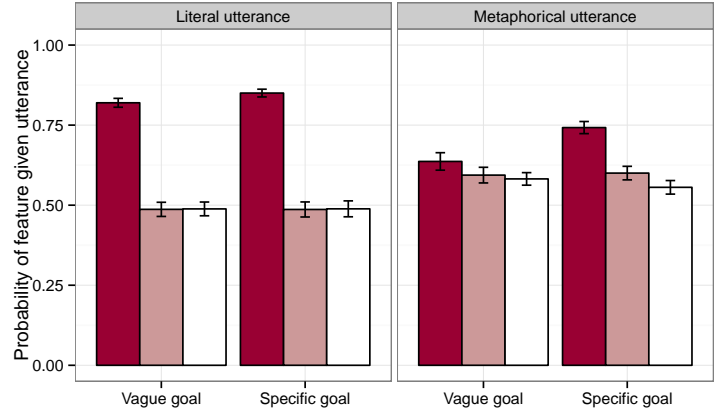


Figure 2: This is a figure.

Table 2: This is a table			
Goal	Utterance	Example question	Example utterance
Vague	Literal	“What is he like?”	“He is scary.”
Specific	Literal	“Is he scary?”	“Yes.”
Vague	Metaphorical	“What is he like?”	“He is a shark.”
Specific	Metaphorical	“Is he scary?”	“He is a shark.”

condition. For each trial, subjects used sliders to indicate the marginal probabilities that the person described has features $f1$, $f2$, and $f3$.

Results For each condition of each scenario, we obtained the average probability ratings for the three features. Figure 2 shows the average ratings for each feature across animal categories given a vague or specific goal and a literal or metaphorical utterance. We see that when the speaker gives a literal statement directly affirming the presence of $f1$, subjects rate $f1$ as significantly more likely than when the speaker gives a metaphorical statement. However, subjects rate $f2$ and $f3$ as significantly more likely when the speaker produces a metaphorical utterance. We also see an effect of the speaker’s communicative goal on the interpretation of metaphorical utterances. Given a specific question about $f1$, subjects interpret the speaker’s metaphorical utterance as being relevant to the question and rates the probability of $f1$ as significantly higher than when the question is vague. On the other hand, the probabilities of $f2$ and $f3$ are not significantly different given a vague communicative goal or a specific goal about $f1$.

Model Evaluation

We used the feature priors $P(\vec{f}|c)$ obtained in Experiment 1B to compute model interpretations of the 32 metaphors. From the behavioral results in Experiment 2, we see evidence that the context set up by a question changes subjects’ interpretation of a metaphor. Our model accounts for this by adjusting the speaker’s prior over communicative goals $P(g|\vec{f})$ un-

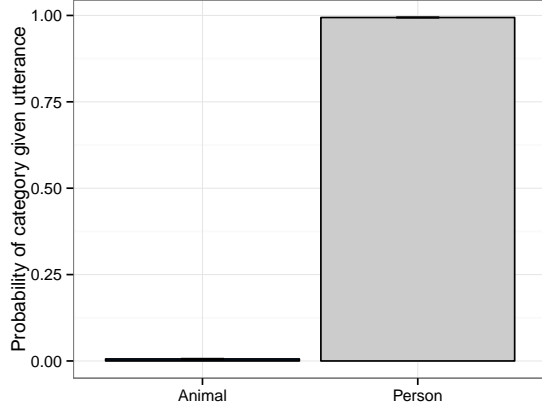


Figure 3: This is a figure.

der different conversational contexts. When a speaker is responding to a vague question, we set the prior distribution for $P(g|\vec{f})$ as uniform. When the speaker is responding to a question specifically about f_1 , we assume that the $P(g_1|\vec{f}) > 0.5$ and equal among $P(g_2|\vec{f}) = P(g_3|\vec{f})$. Fitting the goal prior parameter to data yields a prior of $P(g_1|\vec{f}) = 0.6$ when responding to a specific question about f_1 . We fit the category prior $P(c_a) = 0.01$ and the speaker optimality parameter $\lambda = 3$.

Using these parameters and the model we described, we obtained interpretation probabilities for each of the 32 metaphors under both vague and specific goal conditions. For each metaphor and goal condition, the model produces a joint posterior distribution for $P(c, \vec{f}|u)$. We first show a basic but important qualitative result, which is that the model is able to interpret utterances metaphorically. The pragmatic listener successfully infers that the person described is actually a person and not an animal. Marginal over values of \vec{f} , the probability of category c_p is extremely high ($P(c_p|u) = 0.994$), indicating that the model is able to combine prior knowledge and reasoning about the speaker’s communicative goal to arrive at nonliteral interpretations.

We now turn to the second component of the interpretation, $P(\vec{f}|u)$. Figure X shows the average marginal feature probabilities given a vague or specific goal. The model’s interpretations are shaped by the communicative goal set up by context, where f_1 receives a significantly higher probability when the speaker is responding to a specific question about f_1 . This qualitatively matches behavioral results in Figure X.

To quantitatively evaluate the model’s performance, we correlated model predictions with human interpretations of the metaphorical utterances. Given a metaphorical utterance and a vague or specific question condition, we computed the model’s marginal posterior probabilities for f_1 , f_2 , and f_3 . We then correlate these posterior probabilities with subjects’ probability ratings from Experiment 2. Figure X plots model interpretations for all metaphors, features, and goal condi-

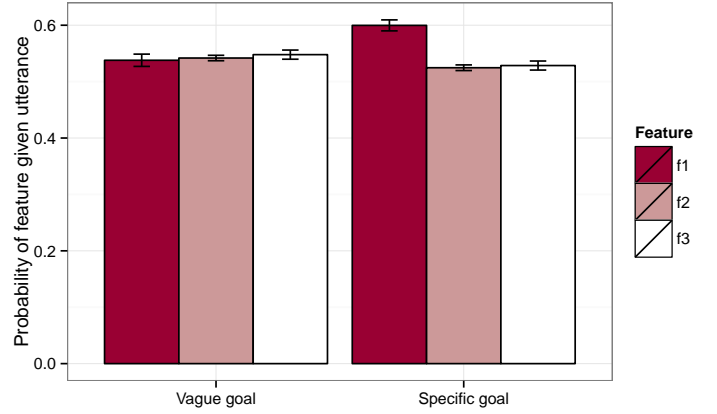


Figure 4: This is a figure.

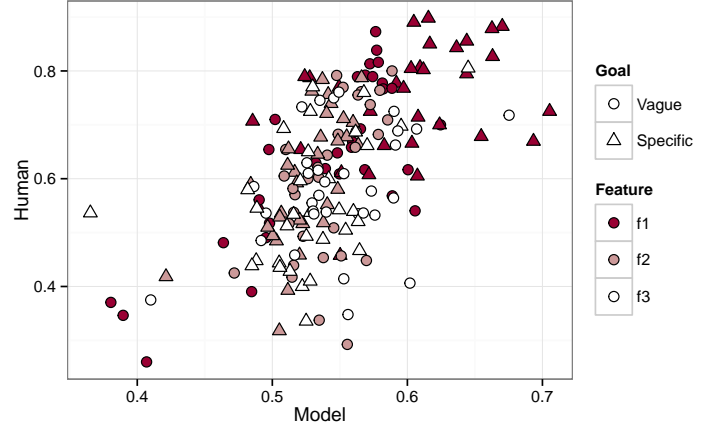


Figure 5: This is a figure.

tions against human judgments. Correlation across the 192 items (32 metaphors \times 3 features \times 2 goal conditions) is 0.6 ($p < 0.001$). The predicted reliability of subjects’ ratings using the Spearman-Brown prediction formula is 0.83, suggesting that our model captures a significant amount of the reliable variance in the behavioral data. In particular, our model does especially well at predicting subject’s judgments of f_1 , which are the most salient features of the animal categories and can be targeted by specific questions in Experiment 2. Correlation between model predictions and human judgments for f_1 is 0.7 ($p < 0.0001$), while the predicted reliability of subjects’ ratings is 0.82.

We now compare our model’s performance to a reasonable baseline model that considers the feature priors for animals and people as well as the conversational context. We constructed a linear regression model that takes the marginal feature priors for the animal category, the marginal feature priors for the person category, and the vague or specific goal as predictors of subjects’ ratings. This model produced an adjusted

$R^2 = 0.19$ ($r = 0.44$), which is a significantly worse fit than our model ($p < 0.0001$ on a Cox test). This suggests that our computational model adequately combines people's prior knowledge about the source and target domains as well as principals of pragmatics to produce metaphor interpretations that closely fit behavioral data.

While our model captures a significant amount of reliable variance, here we perform in-depth analysis of the metaphors and features for which our model performs less well.

Discussion

In this paper we presented a computational model that predicts rich metaphorical interpretations by formalizing principals of communication. Besides going beyond the literal meaning of an utterance to infer non-literal interpretations (i.e. John is a person and not a shark), our model is able to quantitatively predict fine-grained judgments about the person's features (i.e. John is scary, dangerous, and mean). Furthermore, behavioral results show that the interpretation of a metaphor is shaped in part by the conversational context, which our model naturally accounts for by adjusting the priors over conversational goals. Together these results suggest that basic principals of communication may be an important driver of metaphor understanding.

Our model captures several important intuitions about communication, including the importance of common ground between listener and speaker, the communicative goals shaped by the conversational context, and the idea that speakers choose to produce utterances that maximize informativeness about features relevant to their communicative goal. Each of these components inspire research questions that can be further investigated using both our modeling framework and experimental paradigm. For example, are listeners less likely to interpret an utterance metaphorically when there is little common ground between speaker and listener? Does speaking metaphorically allow speakers to respond to specific questions while simultaneously introducing new information that they may wish to communicate? We aim to address these questions in future research to further clarify how components of communication interact to produce metaphorical meaning. We believe that our computational framework for modeling language understanding advances our knowledge of the computational basis of metaphor and other types of nonliteral language, and we hope that it will continue to shed metaphorical light on related questions.

Acknowledgments

Place acknowledgments (including funding information) in a section at the end of the paper.

References

Bergen, B., & Binsted, K. (2003). The cognitive linguistics of scalar humor. *Language, culture, and mind*, 79–92.
 Bergen, L., Goodman, N. D., & Levy, R. (2012). That's what she (could have) said: How alternative utterances affect

language use. In *Proceedings of the thirty-fourth annual conference of the cognitive science society*.
 Blasko, D. G., & Connine, C. M. (1993). Effects of familiarity and aptness on metaphor processing. *Journal of experimental psychology: Learning, memory, and cognition*, 19(2), 295.
 Clark, H. H. (1996). *Using language* (Vol. 4). Cambridge University Press Cambridge.
 Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, 336(6084), 998–998.
 Gibbs Jr, R. W. (2002). A new look at literal meaning in understanding what is said and implicated. *Journal of Pragmatics*, 34(4), 457–486.
 Giora, R. (1997). Understanding figurative and literal language: The graded salience hypothesis. *Cognitive Linguistics*, 8, 183–206.
 Glucksberg, S. (2001). *Understanding figurative language: From metaphors to idioms*. Oxford University Press.
 Glucksberg, S., & Keysar, B. (1990). Understanding metaphorical comparisons: Beyond similarity. *Psychological review*, 97(1), 3.
 Glucksberg, S., & Keysar, B. (1993). How metaphors work.
 Goodman, N. D., & Stuhlmüller, A. (2013). Knowledge and implicature: Modeling language understanding as social cognition. *Topics in cognitive science*, 5(1), 173–184.
 Jäger, G., & Ebert, C. (2009). Pragmatic rationalizability. In *Proceedings of sinn und bedeutung* (Vol. 13, pp. 1–15).
 Kintsch, W., & Bowles, A. R. (2002). Metaphor comprehension: What makes a metaphor difficult to understand? *Metaphor and symbol*, 17(4), 249–262.
 Lakoff, G., et al. (1993). The contemporary theory of metaphor. *Metaphor and thought*, 2, 202–251.
 Lakoff, G., & Turner, M. (2009). *More than cool reason: A field guide to poetic metaphor*. University of Chicago Press.
 Miller, G. A. (1995). Wordnet: A lexical database for english. *COMMUNICATIONS OF THE ACM*, 38, 39–41.
 Ortony, A. (1993). *Metaphor and thought*. Cambridge University Press.
 Ortony, A., Schallert, D. L., Reynolds, R. E., & Antos, S. J. (1978). Interpreting metaphors and idioms: Some effects of context on comprehension. *Journal of Verbal Learning and Verbal Behavior*, 17(4), 465–477.
 Pilkington, A. (2000). *Poetic effects: A relevance theory perspective* (Vol. 75). John Benjamins.
 Roberts, R. M., & Kreuz, R. J. (1994). Why do people use figurative language? *Psychological Science*, 5(3), 159–163.
 Sperber, D., & Wilson, D. (1985). Loose talk. In *Proceedings of the aristotelian society* (Vol. 86, pp. 153–171).
 Sperber, D., & Wilson, D. (2008). A deflationary account of metaphors. *The Cambridge handbook of metaphor and thought*, 84–105.
 Stern, J. J. (2000). *Metaphor in context*. The MIT Press.

- Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction* (Vol. 1) (No. 1). Cambridge Univ Press.
- Tendahl, M., & Gibbs Jr, R. W. (2008). Complementary perspectives on metaphor: Cognitive linguistics and relevance theory. *Journal of Pragmatics*, 40(11), 1823–1864.
- Thibodeau, P. H., & Boroditsky, L. (2011). Metaphors we think with: The role of metaphor in reasoning. *PLoS One*, 6(2), e16782.
- Tourangeau, R., & Sternberg, R. J. (1981). Aptness in metaphor. *Cognitive psychology*, 13(1), 27–55.
- Wilson, D., & Carston, R. (2006). Metaphor, relevance and the emergent property issue. *Mind & Language*, 21(3), 404–433.
- Wilson, D., & Sperber, D. (2002). Relevance theory. *Handbook of pragmatics*.