Pragmatic language interpretation as probabilistic inference

Noah D. Goodman[a,*], Michael C. Frank[a]

[a]*Department of Psychology, Stanford University, 450 Serra Mall, Stanford, CA 94305*

Abstract

Understanding language is more than use of fixed conventions and more than decoding combinatorial structure. Instead, comprehenders make exquisitely sensitive inferences about what utterances mean given their knowledge of the speaker, the language, and the context. Building on developments in game theory and probabilistic modeling, we describe the rational speech act (RSA) model of pragmatic reasoning. RSA and its extensions provide a principled way to formalize inferences about meaning in context. RSA models have been used to make successful quantitative predictions about human behavior in a wide variety of different tasks and situations, and they explain how complex phenomena like hyperbole and vagueness can arise. More generally, they provide a computational framework for integrating linguistic structure, world knowledge, and context in pragmatic language understanding.

> ... one of my avowed aims is to see talking as a special case or variety of purposive, indeed rational, behavior ...
>
> Grice (1975), p. 47

Language is central to the successes of our species; with language we can coordinate our actions, learn from each other, and convey our innermost thoughts. From sounds to syntax, natural languages provide structured methods of combining discrete materials to generate an infinite variety of sentences. Yet this discrete combinatorics does not fully explain how speakers can use language so flexibly to achieve social goals. The interpretation of a particular utterance can itself be almost infinitely variable, depending on factors such as the identity of the speaker, the physical context of its use, and the previous discourse.

*Corresponding author

While the systematization of structural features of language is one of the proudest accomplishments of cognitive science (e.g., Chomsky, 1965; Jackendoff, 2002; Goldberg, 2003), its contextual flexibility – its pragmatics – has been stubbornly difficult to formalize.

Grice (1975) presented an initial framework theory for pragmatic reasoning, positing that speakers are taken to be cooperative, choosing their utterances to convey particular meanings. Gricean listeners then attempt to infer the speaker's intended communicative goal, working backwards from the form of the utterance. This *goal inference* framework for communication has been immensely influential (e.g., Horn, 1984; Sperber & Wilson, 1986; Clark, 1996; Levinson, 2000). But attempts to build on these ideas by providing a specific set of formal principles that allow the derivation of pragmatic inferences have met with difficulty.

For example, the core of Grice's proposal was a set of *conversational maxims* (**see Glossary**; be truthful, relevant, informative, and perspicuous) that could lead to *implicatures* – inferences about speakers' intended meaning – when violated. Formalization of the Gricean notion of implicature using the maxims is difficult (e.g., Hirschberg, 1985), however, and many post-Gricean theories have instead proposed alternative sets of principles (Sperber & Wilson, 1986; Levinson, 2000). An important test of the difficulty of this theoretical project is that the burgeoning experimental psycholinguistic literature attempting to measure pragmatic inference has found these principles only modestly useful (Breheny et al., 2006; Huang & Snedeker, 2009; Noveck & Reboul, 2008). In addition, this sort of informal theory of pragmatics can make only directional, qualitative predictions with respect to experimental data that are typically graded and quantitative.

An alternative strand of Gricean thought has had more success in making contact with data. Grice's core insight was that language use is a form of rational action; thus, technical tools for reasoning about rational action should elucidate linguistic phenomena. Such a goal-directed view of language production has led to the development of engineering systems for natural language generation (Dale & Reiter, 1995) that have in turn been applied as theories of human language production (e.g., Viethen & Dale, 2006). Concurrently, the tools of game-theory – which allow for the characterization of rational actions with respect to defined utilities – have provided a vocabulary for formal descriptions of pragmatic phenomena (e.g., Benz et al., 2006; Jäger, 2008). The recent work we focus on here builds on these developments, combining them with a more detailed view of cognition that arises from the Bayesian cognitive modeling tradition.

Probabilistic, or Bayesian, models have been at the core of a set of recent attempts to understanding the interplay between structured representations and graded or statistical information (Tenenbaum et al., 2011). These models have been an important tool for understanding non-linguistic varieties of rational action, integrating belief understanding with action planning (Baker et al., 2009). A critical feature of these models is that they use the probability calculus to describe inferences under uncertainty. Within formal models of pragmatics, this uncertainty stems from a variety of sources, including uncertainty about speakers' goals and beliefs, uncertainty about the discourse and broader context, and even uncertainty about the meanings of words.

In the remainder of this paper we review the field of probabilistic pragmatics. We begin by describing the *rational speech act* (RSA) model, and the growing body of empirical data supporting its utility in explaining pragmatic reasoning. We next describe extensions to RSA that allow it to be applied to non-literal uses of language like hyperbole, irony and metaphor, to cases of vagueness and ambiguity, and to complex interactions between pragmatics and compositional syntax/semantics. We close with a discussion of broader applications of – and challenges for – probabilistic pragmatics models.

A "rational speech act" model

The "rational speech act" (RSA) model implements a social cognition approach to utterance understanding. At its core, it captures the idea (due to Grice, David Lewis, and others) that speakers are assumed to produce utterances to be helpful yet parsimonious, relative to some particular topic or goal. Listeners then understand utterances by inferring what such a helpful speaker must have meant, given what he said. The first of these basic assumptions is formalized by viewing the speaker as a utility-maximizing agent (where the effort of language production is costly, but communicating information beneficial). The listener then updates his beliefs via Bayesian inference.[1]

The pragmatic listener infers the state of the world, w, using Bayes' rule, given the observation that the speaker chose a particular utterance, u:

$$P_L(w \mid u) \propto P_S(u \mid w)P(w)$$

The key assumption he must make is that the speaker is *approximately rational,* that is, that she has chosen her utterances in proportion to the utility she expects to gain.

$$P_S(u \mid w) \propto exp(\alpha U(u; w))$$

The speaker chooses u from a set of *alternative utterances* (see Outstanding Questions). The parameter α captures the extent to which the speaker maximizes her utility – *how* rational will she be. The basic speaker utility used in RSA captures the social benefit of providing epistemic help to a listener:

[1] For clarity throughout, we use a female pronoun for Alice, the speaker, and a male pronoun for Bob, the listener.

$$U(u;w) = \log P_{\text{Lit}}(w \mid u)$$

This expression measures how certain the listener becomes about the intended world, after hearing the utterance; in order to avoid an infinite recursion, the speaker is assumed to consider a simpler listener, the *literal listener* P_{Lit}. The literal listener again updates his beliefs in accord with Bayesian inference, under the assumption that the literal meaning of the utterance is true:

$$P_{\text{Lit}}(w \mid u) \propto \delta_{\llbracket u \rrbracket(w)} P(w)$$

This definition of the literal listener requires a semantic denotation for each sentence, $\llbracket u \rrbracket$, in which a sentence has the value true or false when applied to a particular state of affairs (i.e. a truth-function on possible worlds). As we describe briefly below, this denotation connects RSA to work in lexical and compositional semantics (Heim & Kratzer, 1998; Dowty et al., 2012).[2]

Consider the scenario in Figure 1, in which speaker and listener share a world of three faces – one with hat and glasses (HG), one with glasses only (G), and one with neither (N); one of these (known to the speaker but not the listener) is the "friend." The speaker says "my friend has glasses," presupposing that there is a single friend. Experimental participants tend to share the intuition that this sentence refers to G and not HG or N (Stiller et al., 2011, 2015).

Under RSA, listener L reasons about S (a simplified internal representation of the speaker), who in turn reasons about Lit (a yet more simplified internal model of the listener). Lit updates his beliefs based on a straightforward denotation: "glasses" applies to both G and HG, while "hat" applies only to HG. Thus $P_{\text{Lit}}(w \mid \text{"hat"})$ places all probability on the friend being HG, while $P_{\text{Lit}}(w \mid \text{"glasses"})$ places equal probability on G and HG (see innermost thought bubbles in Figure 1). The speaker S who intends to communicate that HG is the friend will thus tend to choose the more informative "hat"; but if she intends to communicate that G is the friend, she will use "glasses." Finally, upon hearing "glasses" the listener L infers that this likely refers to G (reflecting the counterfactual that if S had been talking about HG, she would have said "hat" instead).

In the simplest RSA model, as illustrated above, the speaker values providing epistemic help – information – to the listener. But the model can also be extended to create a more sophisticated speaker who is uncertain about the world state, who avoids costly utterances, or who aims to provide relevant information. Connections to other theoretical approaches and aspects of language then

[2]Though traditional Montagovian semantics uses binary truth values, it is also in principle possible in this model to use a semantics with real-valued truth functions, and perhaps even to construct a real-valued semantics from vector-based representations (Monroe & Potts, 2015).

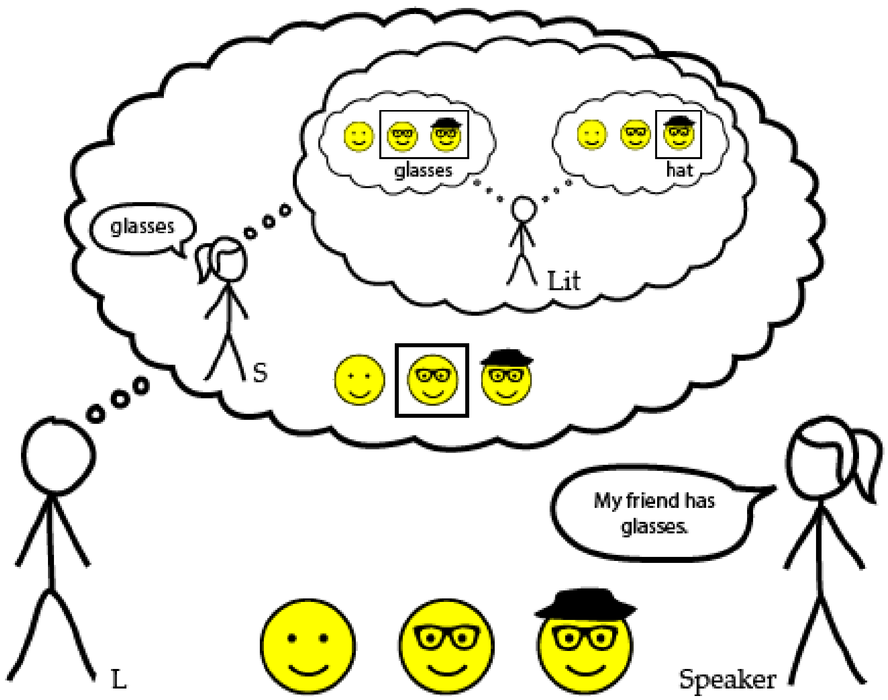

Figure 1: Application of RSA-style reasoning to a signaling game (shown by the three faces along the bottom). Agents are depicted as reasoning recursively about one another's beliefs: listener L reasons about an internal representation of a speaker S, who in turn is modeled as reasoning about a simplified literal listener, Lit. Boxes around targets in the reference game denote interpretations available to a particular agent.

become straightforward. For instance, by modifying the speaker's utility function, we can model the notion of *topic-relevant* information, which connects to linguistic ideas about the "question under discussion" (Roberts, 1996). As a second example, RSA can be combined with the noisy channel approach to language comprehension (Levy, 2008), in order to explain the communicative use of sentence fragments and prosodic stress (Bergen et al., 2016).

In sum, RSA models replace Grice's maxims with a single, utility-theoretic version of the cooperative principle, in which utilities can reflect the communicative and social priorities of a complex, real-world agent.

Empirical support for RSA

The example shown above in Figure 1 is an instance of a signaling game of the type initially introduced by Lewis (1969). Such games are a valuable tool for exploring pragmatic inferences in context, and experiments testing the RSA framework have used games of this type to make quantitative measurements of a variety of different inferences. For example, Frank & Goodman (2012) used a one-shot, web-based paradigm to present participants with geometric shapes in a variety of different configurations. Using a betting paradigm (participants were asked to distribute $100 between response options), these experiments collected separate judgments about what a speaker would say, a listener would interpret, and about baseline expectations for reference (corresponding to the prior $P(w)$). The RSA model showed a tight, parameter-free fit to listeners' aggregate judgements when combined with empirical measurements of the prior distribution.

Although in this initial work RSA was used to simulate the behavior of both speakers and listeners, most subsequent work has focused on the behavior of listeners alone. This work follows the idea that RSA captures listeners' (perhaps optimistic) assumptions about the rational behavior of speakers. Thus, RSA is 'rational' in the sense of *assuming* that speakers are rational; a separate question is how rational speakers actually *are* (**see Box on Producing referring expressions**). In addition, though most research using RSA models has focused on mature language comprehenders, these models have also been the inspiration for a variety of developmental work (**see Box on Children's developing pragmatic competence**).

A variety of other work has replicated and extended the initial findings using similar signaling-game paradigms. In a tight replication of the initial results, Qing & Franke (2015) reproduced the basic findings and explored a set of variants to the initial RSA utility function. And Carstensen et al. (2014) also found that RSA predicted judgments in a communication game using much more complex spatial language stimuli, albeit with somewhat noisier fits. Thus, RSA with an epistemic utility can predict judgments in simple signaling games across variations in both sample and stimulus.

One question raised by this initial work was the level of recursion that best fit human performance. The presentation of RSA given above is stated in terms of a minimal recursion (a listener reasons about speaker, who in turn reasons about a literal listener) but much greater depths of reasoning are in principle possible. The evidence is mixed on whether deeper levels of recursion are commonly seen in language comprehension, however. In a variety of experiments exploring this issue, participants tended to show chance-level performance for signaling systems that required deeper levels of recursion to find unique interpretations (Stiller et al., 2011; Degen & Franke, 2012; Vogel et al., 2013b). More recently, however, Franke & Degen (2016) showed some evidence of deeper recursion for a subpopulation of participants (approx. 15%) in a more complex paradigm, consistent with work on competitive economic games where deeper recursions are sometimes found (Camerer et al., 2004). This heterogeneity – and its dependence on individual and contextual differences – is an interesting topic for future work.

A number of other studies have tested RSA with more elaborated utility functions (**see Box on Refinements to the speakers utility**). For instance, a speaker might be expected to produce a less informative utterance when the more informative one is much harder to say. This tendency can be formalized by including a cost term in the speaker's utility; with this modification, RSA predicts the impact of production costs on interpretations of the listener. Exploring this extension, Bergen et al. (2012) showed that participants in a reference game are indeed sensitive to the cost, in dollars, of alternative message choices. And Degen et al. (2013) tested the effect of production difficulty by manipulating how quickly the speaker could type on an on-screen keyboard; participants' interpretations reflected this difficulty as predicted. Additional work has used proxies for production cost such as number of words and their frequencies in explorations of phenomena such as negation and nominal reference (Nordmeyer & Frank, 2014; Graf et al., 2016).

Finally, in addition to ad-hoc signaling systems, RSA provides a way to describe reasoning about classic linguistic implicatures. Perhaps the best-studied of these is the scalar implicature that "some of the letters had checks inside" implicates that *not all* did. Goodman & Stuhlmüller (2013) measured participants' judgments about the interpretations of quantifiers and number words in exactly this situation and found that these judgments were well-predicted by RSA. In addition, a critical feature of this study was the inclusion of an epistemic manipulation (e.g., that some of the letters had not been opened yet). By using expected informativity to account for the speaker's limited perceptual access, the model was able to predict differing patterns of listener judgments based on different levels of speaker uncertainty. These empirical findings are congruent with other recent demonstrations of the importance of epistemic reasoning in pragmatic implicature (Bergen et al., 2012; Breheny et al., 2013, e.g.,). They also highlight the way the RSA framework provides a (non-modular) theory for interactions between language and non-linguistic cognition. We next turn to a variety of other extensions to the basic RSA model that explore additional interactions.

Uncertainty about the speaker: Joint reasoning

In the basic RSA model, the listener has a specific model in mind of how a speaker will behave. But what should a listener do if he is not sure what speaker model is appropriate? Recent work has answered this question by positing a joint inference: Which kind of speaker am I interacting with and what is the world like, given the utterance I heard? Formally this *uncertain* RSA (or uRSA) framework requires only a small change:

$$P_L(w, s \mid u) \propto P_S(u \mid w, s)P(s)P(w)$$

where the new variable s parametrizes different speaker types. In practice, s can refer to any factor that might influence the speaker's behavior, including uncertainty about conversational topic, word meanings, background knowledge, or general discourse context. This modification allows uRSA to capture a much wider variety of linguistic phenomena; intuitively, an uRSA listener is a more realistic cognitive agent than the RSA listener, who was restricted to the specifics of a very particular context and goal. To illustrate this intuition, we provide three examples of phenomena captured by uRSA (but not by basic RSA): non-literal language, vagueness, and embedded implicatures.

Non-literal or figurative language – utterances that are easily interpreted but not "actually true" – poses a problem for nearly all formal models of language understanding. How can tropes like hyperbole, sarcasm, and metaphor be interpreted, and why are they used? Under uRSA, these uses can be described as arising from uncertainty about the topic of conversation. If the speaker is expected to provide information relevant for a particular topic, the pragmatic listener will only update his beliefs along this topical dimension. Within uRSA, the interaction between uncertainty about the speaker's intended topic and her intended meaning about that topic can drive complex interpretations.

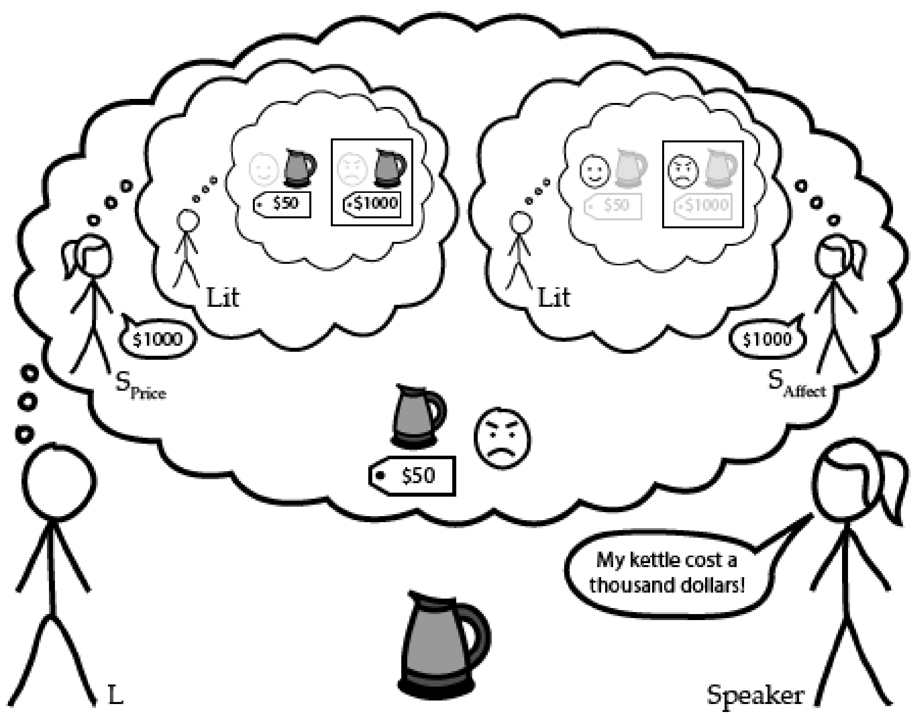

Figure 2: Uncertain RSA-style reasoning applied to hyperbole. Listener L reasons jointly about the price of the item and the speaker's affect. In doing so he considers two speakers, one who is primarily interested in conveying her affective response to the kettle, and one who is primarily interested in conveying the actual price. Each of these speakers is modeled as reasoning about literal listeners who focus on different aspects of the situation.

Kao et al. (2014b) investigated hyperbolic utterances such as “the electric kettle cost $1,000” In this example, the number $1,000 can interpreted as conveying information about the speaker’s affect, not the actual price, in part because one thousand dollars is an implausibly high price for a kettle. As shown in Figure 2, the uRSA model captures this intuition by positing that the topic of the speaker’s utterance may be the actual price of the kettle, the speaker’s opinion about the price, or some combination of the two. Since the listener does not know the topic, he jointly infers it together with the likely true price of the kettle and the speaker’s affect. When the uttered price is implausible this joint inference will yield a non-standard topic and hence a non-literal interpretation.

By extending the space of affect to include both valence and arousal, Kao & Goodman (2015) showed that the same model predicts verbal irony. And a similar approach has been suggested for simple metaphors (Kao et al., 2014a), such as “John is a shark.” Here the potential topics include not affect, but features of the target, such as how vicious John is and how likely he is to swim underwater. In each of these cases of figurative language, the uRSA model accounts for almost all of the explainable variance in human interpretations, a striking result considering the complexity and subtlety of these phenomena.

Linguistic descriptions, especially adjectives, are both context-sensitive and vague. Providing precise definitions for words like *expensive* or *tall* has been a persistent challenge for philosophers and semanticists (Williamson, 2002). uRSA models address this challenge by assuming that word meanings can differ between speakers and contexts, and and that these meanings themselves can be a subject for inference. In the case of scalar adjectives like *tall*, the uncertainty is over the threshold required: what height is required before an object counts as tall? Under the uRSA model, judgments about meaning take into account two conflicting pressures: a stricter threshold for tallness makes the term *tall* more informative, but the term must be loose enough to be true. By negotiating this balance, uRSA accounts for three key phenomena of vague adjectives (Lassiter & Goodman, 2015): the inferred meaning depends on the class (tall for a tree vs. tall for a person), there are borderline cases, and the interpretations are subject to a *sorites* paradox (no single minimal increment in height will make you tall, but enough increments will). The process of reasoning about meaning modeled by uRSA might even interact with learning processes to produce more long-lasting inferences about word meaning, leading to language change (**see Box on Language use and language change**).

Finally, this uRSA approach allows for progress on an important puzzle in recent discussions of pragmatic inference: embedded implicature (Geurts & Pouscoulous, 2009; Chemla & Spector, 2011). Embedded implicatures occur when quantifiers are nested within one another, as in sentences like “Exactly one letter is connected with some of its circles.” In these cases, some experimental evidence suggests that participants access the interpretation that one letter is connected with some *but not all* of its circles, an interpretation that standard Gricean theories cannot generate (Chemla & Spector, 2011). Potts et al. (2015)

replicated these interpretations in a series of large-scale experiments and confirmed that basic RSA models could not capture them. An implementation of uRSA that operates over fully-compositional semantic systems (Bergen et al., 2016; Potts et al., 2015) showed a good fit to the overall pattern of data, however, supporting the idea that uRSA is a fruitful way to incorporate pragmatic reasoning into compositional systems.

Conclusions

Context-dependence is one of the core features of natural language. Yet because of the informal nature of theorizing about this context-dependence, pragmatics has often been treated as a theoretical "wastebasket," in which unexplained phenomena are hidden (Bar-Hillel, 1971). Countering this trend, new formal theories of pragmatics make quantitative predictions about a wide variety of phenomena that have previously been considered too difficult to operationalize. These include implicature, vagueness, non-literal language, and the myriad other cases where linguistic meaning is changed by context.

The key tool in this work is the Rational Speech Act model, which builds upon and synthesizes a number of formal traditions in the study of human inference, from game theory to models of human reasoning. The RSA approach builds on existing work on semantic representation – using a compositional semantics *a la* Montague (Dowty et al., 2012) – and contributes back to semantics by providing a specific mechanism by which underspecified meanings become precise, in context.

The RSA framework is a *computational* level description of the language user's competence (Marr, 1982). There are many possible ways a cognitive agent could implement RSA at the algorithmic level. These alternatives must be evaluated for their ability to capture the process of human language understanding and production (Degen & Tanenhaus, 2015; Nordmeyer & Frank, 2014). Yet even in its current form, RSA potentially suggests a new take on pragmatic language processing – one that is consistent with recent empirical findings. In Gricean analyses, a violation of a maxim leads to reasoning to "repair" the interpretation. Many theorists have inferred from this idea of repair that pragmatic inferences should be slow (because they depend on full semantic interpretation) and optional (because they only happen after a violation). These assumptions have been challenged both theoretically and empirically (e.g., Levinson, 2000; Grodner et al., 2010). In contrast, RSA-style reasoning makes pragmatic inferences a fundamental part of language comprehension, in which the ultimate goal of all interpretation is to settle on the intended meaning, given both the literal semantics of the utterance and the broader pragmatic context. In this way, RSA is consistent with modern psycholinguistic theories that emphasize interactive, incremental processing (Degen & Tanenhaus, 2015).

Future extensions of RSA will likely include more complex worlds; pragmatic alternatives with compositional structure; more sophisticated, multi-part discourses; and utility structures that better take into account the complexities of social interaction. On the practical side, computing the predictions of RSA models can become prohibitive when the number of world states or utterances grows large. Further development of algorithms to implement RSA are needed. These developments may go together with new algorithms for learning aspects of the underlying semantics, which will open up new applications for the RSA approach in computational linguistics and artificial intelligence (Golland et al., 2010; Vogel et al., 2013a; Monroe & Potts, 2015; Andreas & Klein, 2016).

The work outlined in this review represents steps towards a comprehensive, formal theory of language understanding in context. While much further work will be required, RSA models and their uRSA extensions have proven to be useful tools for explaining both qualitative and quantitative empirical data across a wide range of tasks and contexts. Language is central to the human experience. We hope our work sheds light on how its structure and systematicity can still give rise to such an astonishingly flexible communication system.

Acknowledgements

We gratefully acknowledge funding from a James S. McDonnell Foundation Scholar Award to Goodman, ONR Grants N00014-13-1-0788 and #N00014-13-1-0287 to Goodman and Frank, and NSF BCS Grant #1456077 to Frank.

References

Andreas, J., & Klein, D. (2016). Reasoning about pragmatics with neural listeners and speakers. *arXiv preprint arXiv:1604.00562*, .

Baker, C. L., Saxe, R., & Tenenbaum, J. B. (2009). Action understanding as inverse planning. *Cognition*, *113*, 329–349.

Bar-Hillel, Y. (1971). Out of the pragmatic wastebasket. *Linguistic inquiry*, *2*, 401–407.

Barner, D., Brooks, N., & Bale, A. (2011). Accessing the unsaid: The role of scalar alternatives in children's pragmatic inference. *Cognition*, *118*, 84 – 93. doi:10.1016/j.cognition.2010.10.010.

Baumann, P., Clark, B., & Kaufmann, S. (2014). Overspecification and the cost of pragmatic reasoning about referring expressions. *Proceedings of CogSci. Austin, TX: Cognitive Science Society*, .

Benz, A., Jäger, G., & Van Rooij, R. (2006). An introduction to game theory for linguists. In *Game theory and pragmatics* (pp. 1–82). Springer.

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 34th Annual Conference of the Cognitive Science Society*.

Bergen, L., Levy, R., & Goodman, N. D. (2016). Pragmatic reasoning through semantic inference. *Semantics & Pragmatics*, .

Breheny, R., Ferguson, H. J., & Katsos, N. (2013). Taking the epistemic step: Toward a model of on-line access to conversational implicatures. *Cognition*, *126*, 423–440.

Breheny, R., Katsos, N., & Williams, J. (2006). Are generalised scalar implicatures generated by default? an on-line investigation into the role of context in generating pragmatic inferences. *Cognition*, *100*, 434–463.

Camerer, C. F., Ho, T.-H., & Chong, J.-K. (2004). A cognitive hierarchy model of games. *The Quarterly Journal of Economics*, (pp. 861–898).

Carstensen, A., Kon, E., & Regier, T. (2014). Testing a rational account of pragmatic reasoning: The case of spatial language. In *Proceedings of the 36 th Annual Meeting of the Cognitive Science Society. Austin, TX: Cognitive Science Society* (pp. 2009–2013).

Chemla, E., & Spector, B. (2011). Experimental evidence for embedded scalar implicatures. *Journal of semantics*, (p. ffq023).

Chomsky, N. (1965). *Syntactic structures*. Walter de Gruyter.

Clark, E. V., & Amaral, P. M. (2010). Children build on pragmatic information in language acquisition. *Language and Linguistics Compass*, *4*, 445–457.

Clark, H. (1996). *Using Language*. Cambridge, UK: Cambridge University Press.

Dale, R., & Reiter, E. (1995). Computational interpretations of the gricean maxims in the generation of referring expressions. *Cognitive science*, *19*, 233–263.

Degen, J., & Franke, M. (2012). Optimal reasoning about referential expressions. *Proceedings of SemDIAL*, .

Degen, J., Franke, M., & Jäger, G. (2013). Cost-based pragmatic inference about referential expressions. In *Proceedings of the 35th annual conference of the cognitive science society* (pp. 376–381).

Degen, J., & Tanenhaus, M. K. (2015). Processing scalar implicature: A constraint-based approach. *Cognitive science*, *39*, 667–710.

Dowty, D. R., Wall, R., & Peters, S. (2012). *Introduction to Montague semantics* volume 11. Springer Science & Business Media.

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

Frank, M. C., & Goodman, N. D. (2014). Inferring word meanings by assuming that speakers are informative. *Cognitive Psychology*, *75*, 80–96. URL: `http://langcog.stanford.edu/papers/FG-cogpsych2014.pdf`.

Franke, M., & Degen, J. (2016). Reasoning in reference games: Individual-vs. population-level probabilistic modeling. *PloS one*, *11*, e0154854.

Gatt, A., van Gompel, R. P., van Deemter, K., & Krahmer, E. (2013). Are we bayesian referring expression generators. In *Proceedings of CogSci*. volume 35.

Geurts, B., & Pouscoulous, N. (2009). Embedded implicatures?!? *Semantics and pragmatics*, *2*, 4–1.

Goldberg, A. E. (2003). Constructions: a new theoretical approach to language. *Trends in cognitive sciences*, *7*, 219–224.

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). Association for Computational Linguistics.

Goodman, N. D., & Stuhlmüller, A. (2013). Knowledge and implicature: Modeling language understanding as social cognition. *Topics in Cognitive Science*, *5*, 173–184.

Graf, C., Hawkins, R. X. D., & Goodman, N. D. (2016). Animal, dog, or dalmatian? Level of abstraction in nominal referring expressions. In *Proceedings of the 38th Annual Conference of the Cognitive Science Society*.

Grice, H. (1975). Logic and conversation. *1975*, (pp. 41–58).

Grodner, D. J., Klein, N. M., Carbary, K. M., & Tanenhaus, M. K. (2010). Some, and possibly all, scalar inferences are not delayed: Evidence for immediate pragmatic enrichment. *Cognition*, *116*, 42–55.

Heim, I., & Kratzer, A. (1998). *Semantics in generative grammar* volume 13. Blackwell Oxford.

Hirschberg, J. L. B. (1985). *A theory of scalar implicature*. University of Pennsylvania.

Horn, L. R. (1984). Toward a new taxonomy for pragmatic inference: Q-based and R-based implicature. In D. Schiffrin (Ed.), *Meaning, Form, and Use in Context: Linguistic Applications* (pp. 11–42). Washington, D.C.: Georgetown University Press.

Horowitz, A. C., & Frank, M. C. (2016). Children's pragmatic inferences as a route for learning about the world. *Child Development*, *87*, 807–819.

Huang, Y. T., & Snedeker, J. (2009). Online interpretation of scalar quantifiers: Insight into the semantics–pragmatics interface. *Cognitive psychology*, *58*, 376–415.

Jackendoff, R. S. (2002). *Foundations of langage: Brain, meaning, grammar, evolution*. Oxford, UK: Oxford University Press.

Jäger, G. (2008). Applications of game theory in linguistics. *Language and Linguistics compass*, *2*, 406–421.

Kao, J. T., Bergen, L., & Goodman, N. D. (2014a). Formalizing the pragmatics of metaphor understanding. In *Proceedings of the 36th annual meeting of the Cognitive Science Society* (pp. 719–724).

Kao, J. T., & Goodman, N. D. (2015). Let's talk (ironically) about the weather: Modeling verbal irony. In *Proceedings of the 37th Annual Conference of the Cognitive Science Society*.

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

Katsos, N., & Bishop, D. (2011). Pragmatic tolerance: Implications for the acquisition of informativeness and implicature. *Cognition*, *120*, 67–81. doi:10.1016/j.cognition.2011.02.015.

Kemp, C., & Regier, T. (2012). Kinship categories across languages reflect general communicative principles. *Science*, *336*, 1049–1054.

Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition*, *89*, 25–41.

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences*, *105*, 10681–10686.

Kirby, S., Tamariz, M., Cornish, H., & Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition*, *141*, 87–102.

Lane, L. W., Groisman, M., & Ferreira, V. S. (2006). Don't talk about pink elephants! speakers' control over leaking private information during language production. *Psychological science*, *17*, 273–277.

Lassiter, D., & Goodman, N. D. (2015). Adjectival vagueness in a bayesian model of interpretation. *Synthese*, (pp. 1–36).

Levelt, W. J. (1993). *Speaking: From intention to articulation* volume 1. MIT press.

Levinson, S. (2000). *Presumptive meanings: The theory of generalized conversational implicature*. MIT Press.

Levy, R. (2008). A noisy-channel model of rational human sentence comprehension under uncertain input. In *Proceedings of the conference on empirical methods in natural language processing* (pp. 234–243). Association for Computational Linguistics.

Lewis, D. (1969). *Convention: A philosophical study*. John Wiley & Sons.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. New York, NY: Henry Holt and Co.

Monroe, W., & Potts, C. (2015). Learning in the rational speech acts model. *arXiv preprint*, .

Nordmeyer, A. E., & Frank, M. C. (2014). A pragmatic account of the processing of negative sentences. In *Proceedings of the 36th Annual Meeting of the Cognitive Science Society*.

Noveck, I. (2001). When children are more logical than adults: experimental investigations of scalar implicature. *Cognition*, *78*, 165–188. doi:10.1016/S0010-0277(00)00114-1.

Noveck, I. A., & Reboul, A. (2008). Experimental pragmatics: A gricean turn in the study of language. *Trends in Cognitive Sciences*, *12*, 425–431.

O'Neill, D. K., & Topolevec, J. C. (2001). Two-year-old children's sensitivity to the referential (in) efficacy of their own pointing gestures. *Journal of Child Language*, *28*, 1.

Papafragou, A., & Musolino, J. (2003). Scalar implicatures: experiments at the semantics-pragmatics interface. *Cognition*, *86*, 253–282. doi:10.1016/S0010-0277(02)00179-8.

Perfors, A., & Navarro, D. J. (2014). Language evolution can be shaped by the structure of the world. *Cognitive science*, *38*, 775–793.

Piantadosi, S. T., Tily, H., & Gibson, E. (2012). The communicative function of ambiguity in language. *Cognition*, *122*, 280–291.

Potts, C., Lassiter, D., Levy, R., & Frank, M. C. (2015). Embedded implicatures as pragmatic inferences under compositional lexical uncertainty. *Journal of Semantics*, (p. ffv012).

Qing, C., & Franke, M. (2015). Variations on a bayesian theme: Comparing bayesian models of referential reasoning. In *Bayesian natural language semantics and pragmatics* (pp. 201–220). Springer.

Regier, T., Kay, P., & Khetarpal, N. (2007). Color naming reflects optimal partitions of color space. *Proceedings of the National Academy of Sciences*, *104*, 1436–1441.

Roberts, C. (1996). Information structure in discourse: Towards an integrated formal theory of pragmatics. *Working Papers in Linguistics-Ohio State University Department of Linguistics*, (pp. 91–136).

Smith, N. J., Goodman, N., & Frank, M. (2013). Learning and using language via recursive pragmatic reasoning about other agents. In *Advances in Neural Information Processing Systems*. URL: http://langcog.stanford.edu/papers/SGF-NIPS2013.pdf.

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and Cognition*. Oxford, UK: Blackwell Publishers.

Stiller, A., Goodman, N. D., & Frank, M. C. (2011). Ad-hoc scalar implicature in adults and children. In *Proceedings of the 33rd annual meeting of the Cognitive Science Society* (pp. 2134–2139). Cognitive Science Society Austin, TX.

Stiller, A. J., Goodman, N. D., & Frank, M. C. (2015). Ad-hoc implicature in preschool children. *Language Learning and Development*, *11*, 176–190.

Viethen, J., & Dale, R. (2006). Algorithms for generating referring expressions: do they do what people do? In *Proceedings of the Fourth International Natural Language Generation Conference* (pp. 63–70). Association for Computational Linguistics.

Vogel, A., Bodoia, M., Potts, C., & Jurafsky, D. (2013a). Emergence of gricean maxims from multi-agent decision theory. In *HLT-NAACL* (pp. 1072–1081).

Vogel, A., Potts, C., & Jurafsky, D. (2013b). Implicatures and nested beliefs in approximate decentralized-pomdps. In *ACL (2)* (pp. 74–80). Citeseer.

Vouloumanos, A., Onishi, K., & Pogue, A. (2012). Twelve-month-old infants recognize that speech can communicate unobservable intentions. *Proceedings of the National Academy of Sciences*, *109*, 12933–12937.

Williamson, T. (2002). *Vagueness*. Routledge.

Xu, Y., & Regier, T. (2014). Numeral systems across languages support efficient communication: From approximate numerosity to recursion. *Proceedings of the 36th Annual Meeting of the Cognitive Science Society*, .

Yoon, E. J., Tessler, M. H., & Frank, M. C. (2016). Talking with tact: Polite language as a balance between kindness and informativity. In *Proceedings of the 38th Annual Conference of the Cognitive Science Society*.

Trends

- Rational speech act (RSA) models provide a quantitative framework to capture intuitions about pragmatic reasoning in language understanding.
- Extensions to RSA that allow for reasoning about the speaker – for instance, her goals and word usage – can capture many otherwise puzzling phenomena including vagueness, embedded implicatures, hyperbole, irony, and metaphor.
- The RSA framework can inform psycholinguistic processing experiments, linguistic theory, and scalable natural language processing models.

Box: Outstanding questions

- *Model details.* How deeply do human comprehenders recurse when reasoning about others' intentions? Is depth of reasoning constant, or does it vary across situations? How are alternative utterances computed?
- *Linguistic goals.* How do "Gricean" utilities – the drive to be informative yet succinct – relate to other social goals like conveying affect, or establishing relationships? How do cooperative and competitive goals mix in language use?
- *Dialogue.* How can RSA be used to model the evolution over the course of a conversation of a partners' utilities, possible goals, and the context more broadly?
- *Learning and language change.* How do pragmatic language understanding and language learning interact? How and when does pragmatic language use lead to diachronic language change?
- *Algorithmic challenges.* Given the potential complexity of recursive pragmatic computations, how is language processed so quickly? How can RSA models be "scaled up" for natural language processing tasks?

Box: Refinements to the speaker's utility

The notion of the speaker's utility – what is rewarding for a speaker – is central to the RSA approach. The basic RSA model captures the speaker's need to be *informative* to a listener: $U(u; w) = \log P_{\text{Lit}}(w \mid u)$. Different utilities lead to different kinds of speakers, which in turn lead to different interpretations by the pragmatic listener. Several utility refinements (and their combinations) have been considered in recent work:

- Utterance cost: In order to capture a tendency of speakers to be parsimonious we can simply add a cost term: $U(u; w) = \log P_{\text{Lit}}(w \mid u) + cost(u)$. The cost may reflect actual production cost (such as number of words) or proxies such as word frequency. This extension yields effects similar to Grice's maxim of manner (Bergen et al., 2016).

- Speaker uncertainty: When the speaker does not have full knowledge of the world she should choose an utterance according to *expected utility*: $U(u; k) = \mathbb{E}_{P(w|k)}[U(u; w)]$, where k summarizes the speaker's knowledge or observations. This extension correctly predicts interactions between a speaker's knowledge and a listener's interpretations (Goodman & Stuhlmüller, 2013).

- Topic relevance: While it may be highly informative to provide detailed descriptions, such detail is not always *relevant.* Relevance can be captured by introducing a topic of conversation (sometimes known as a *Question Under Discussion,* Roberts, 1996) and adjusting the epistemic utility to reflect only information about this topic: $U(u; w, t) = \log \sum_{w' \text{ s.t. } t(w')=t(w)} P_{\text{Lit}}(w' \mid u)$, where the function t projects from complete worlds to the relevant aspects (Kao et al., 2014b).

- Other social goals: Language is often used not just to inform, but also to flirt, insult, comfort, and to pursue myriad other social goals. For example, non-informational utilities, e.g. utility directed towards kindness, can produce behaviors that appear polite (Yoon et al., 2016).

Box: Children's developing pragmatic competence

From a very early age, children are oriented towards communication, understanding the function of language for information transfer and repurposing their limited linguistic means to achieve a wide variety of ends (Vouloumanos et al., 2012; Clark & Amaral, 2010). In light of this general early orientation, the literature on pragmatic development specifically has been puzzling: older children very reliably fail to make scalar implicatures under a range of circumstances (Noveck, 2001). In one striking example, five-year-olds endorsed the statement that "some of the horses jumped over the fence" even when three out of three of a set of horses had made the jump (Papafragou & Musolino, 2003).

What explains this disconnect between early communicative successes and later pragmatic failures? Early theorizing suggested that younger children might not be able to make pragmatic computations, but this hypothesis appears unlikely given evidence that younger children are sensitive to the pragmatic informativeness of messages (Katsos & Bishop, 2011; O'Neill & Topolevec, 2001). More recently, theorists have proposed that apparent difficulties with pragmatic implicatures may have resulted from children's inadequate knowledge of the specific lexical alternatives that compete in pragmatic inferences (e.g., not understanding that "some" contrasts with "all," leading to the implicature "some but not all") rather than difficulty with pragmatic computations more generally (Barner et al., 2011).

Our work supports this "alternatives hypothesis" by providing evidence for a range of implicature computations in young children. For example, in a purely referential (or "ad hoc") task that did not use quantifiers or other complex language, three-year-old children showed signs of successful implicature computations (Stiller et al., 2015). And children in the same age range were able to use an implicature to guess the meaning of a novel word (Frank & Goodman, 2014) or a novel context (Horowitz & Frank, 2016). These findings support the idea that even young children are able to make flexible pragmatic inferences in simple referential scenarios, and are consistent with the application of RSA-style reasoning. Future research will be required, however, to test whether RSA (or some capacity-limited modification) could make quantitative predictions about pragmatic development.

Box: Producing referring expressions

RSA stands for the "rational speech act" model, indicating that listeners idealize speakers as rational. Are speakers in fact rational in a meaningful way? And if so, how can this conclusion be integrated with the large body of evidence indicating that speakers are egocentric, error prone, and subject to idiosyncratic production preferences (Keysar et al., 2003; Lane et al., 2006; Gatt et al., 2013)?

Although our initial studies collected judgements about language production in extremely restricted tasks (Frank & Goodman, 2012), most recent work using the RSA model has focused on modeling listeners' judgements, rather than speakers' productions. One reason for this choice is that often the most interesting pragmatic inferences come about when speakers are not maximally informative. For example, in the signaling game shown in Figure 1, helpful speakers will often overspecify and say "glasses and no hat" (Baumann et al., 2014). But this seemingly non-pragmatic response may in fact be a reasonable response to uncertainty about whether a conversational partner will in fact draw the desired implicature. More generally, speakers' production choices are a promising area for future research using RSA models with a broader range of utility functions (see Box) and various sources of potential miscommunication (a topic of ongoing research).

Nevertheless, it is clear that in their natural behavior, speakers make production decisions under time pressure and a variety of cognitive demands (Levelt, 1993). Integrating these demands with the predictions of utility-theoretic models should be an important challenge for future work.

Box: Language use and language change

The pragmatic processes described by RSA models occur in the moment of communication, but can have a set of effects that ripple out through language as a whole. The construal of an individual communication event can influence learning processes (see Box), which in turn can lead to systematic changes in word meanings (Smith et al., 2013). Words that are too narrow in their denotation can be pragmatically extended (Kao et al., 2014b), while words that are too broad can be narrowed via implicature. Over time, word meanings may converge to the appropriate level of ambiguity to enable efficient communication (Piantadosi et al., 2012). In this sense, in-the-moment pragmatic interpretation may bootstrap long term language change.

The processes of change that promote efficient communication have been explored extensively within the iterated learning paradigm. For example, iterated transmission of an arbitrary communication system can lead to the emergence of semantic regularities (Kirby et al., 2008). This framework can also be used to express the competing pressures of learnability and communication. When languages are selected only to be learnable, they often become degenerate, including only a single word (Perfors & Navarro, 2014). But when they include a countervailing pragmatic pressure, which can be modeled via RSA, expressive and compositional languages can emerge (Kirby et al., 2015).

If pressures for efficient communication lead to language change, then these pressures should be visible in the lexicons of human languages. A recent body of evidence suggests that they are. Regier and colleagues have argued that the typological distribution of languages in particular semantic domains reflects the range of optimal communication systems (e.g., Regier et al., 2007; Kemp & Regier, 2012; Xu & Regier, 2014). An important future direction is to understand whether this typological distribution is predicted to arise from iterated learning with a population of RSA language users.

Glossary

- **Rational speech act model (RSA)**. A class of probabilistic model that assumes that language comprehension in context arises via a process of recursive reasoning about what speakers would have said, given a set of communicative goals.
- **Implicature**. An inference about the meaning of an utterance in context that goes beyond its literal semantics. Implicatures are typically *cancellable* in that they can be contradicted, as in "Some of the students passed the exam; indeed, all of them did."
- **Conversational maxims**. A set of principles described by Grice (1975) as a theory of how listeners reason about speakers' intended meaning to arrive at pragmatic implicatures.

- **Uncertain RSA models (uRSA)**. A specific extension of RSA models that allow for joint inferences about both the speaker's intended meaning and other aspects of the interaction, such as the topic, the context, or even the meanings of particular words.