Implicatures and Nested Beliefs in Approximate Decentralized-POMDPs

Adam Vogel, Christopher Potts, and **Dan Jurafsky**
Stanford University
Stanford, CA, USA
{acvogel,cgpotts,jurafsky}@stanford.edu

Abstract

Conversational implicatures involve reasoning about multiply nested belief structures. This complexity poses significant challenges for computational models of conversation and cognition. We show that agents in the multi-agent Decentralized-POMDP reach implicature-rich interpretations simply as a by-product of the way they reason about each other to maximize joint utility. Our simulations involve a reference game of the sort studied in psychology and linguistics as well as a dynamic, interactional scenario involving implemented artificial agents.

1 Introduction

Gricean conversational implicatures (Grice, 1975) are inferences that listeners make in order to reconcile the speaker's linguistic behavior with the assumption that the speaker is cooperative. As Grice conceived of them, implicatures crucially involve reasoning about multiply-nested belief structures: roughly, for p to count as an implicature, the speaker must believe that the listener will infer that the speaker believes p. This complexity makes implicatures an important testing ground for models of conversation and cognition.

Implicatures have received considerable attention in the context of simple reference games in which the listener uses the speaker's utterance to try to identify the speaker's intended referent (Rosenberg and Cohen, 1964; Clark and Wilkes-Gibbs, 1986; Dale and Reiter, 1995; DeVault and Stone, 2007; Krahmer and van Deemter, 2012). Many implicature patterns can be embedded in these games using specific combinations of potential referents and message sets. The paradigm has proven fruitful not only for evaluating computational models (Golland et al., 2010; Degen and Franke, 2012; Frank and Goodman, 2012; Rohde et al., 2012; Bergen et al., 2012) but also for studying children's pragmatic abilities without implicitly assuming they have mastered challenging linguistic structures (Stiller et al., 2011).

In this paper, we extend these results beyond simple reference games to full decision-problems in which the agents reason about language and action together over time. To do this, we use the Decentralized Partially Observable Markov Decision Process (Dec-POMDP) to implement agents that are capable of manipulating the multiply-nested belief structures required for implicature calculation. Optimal decision making in Dec-POMDPs is NEXP complete, so we employ the single-agent POMDP approximation of Vogel et al. (2013). We show that agents in the Dec-POMDP reach implicature-rich interpretations simply as a by-product of the way they reason about each other to maximize joint utility. Our simulations involve a reference game and a dynamic, interactional scenario involving implemented artificial agents.

2 Decision-Theoretic Communication

The Decentralized Partially Observable Markov Decision Process (Dec-POMDP) (Bernstein et al., 2002) is a multi-agent generalization of the POMDP, where agents act to maximize a shared utility function. Formally, a Dec-POMDP consists of a tuple $(S, A, O, R, T, \Omega, b_0, \gamma)$. S is a finite set of states, A is the set of actions, O is the set of observations, and $T(s'|a_1, a_2, s)$ is the *transition distribution* which determines what effect the joint action (a_1, a_2) has on the state of the world. The true state $s \in S$ is not observable to the agents, who must utilize observations $o \in O$, which are emitted after each action according to the *observation distribution* $\Omega(o_1, o_2|s', a)$. The *reward function* $R(s, a_1, a_2)$ represents the goal of the agents, who act to maximize expected reward. Lastly, $b_0 \in \Delta(S)$ is the initial belief state and

Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, pages 74–80,
Sofia, Bulgaria, August 4-9 2013.

$\gamma \in [0, 1)$ is the discount factor.

The true state of the world $s \in S$ is not observable to either agent. In single-agent POMDPs, agents maintain a *belief state* $b(s) \in \Delta(S)$, which is a distribution over states. Agents acting in Dec-POMDPs must take into account not only their beliefs about the state of the world, but also the beliefs of their partners, leading to *nested* belief states. In the model presented here, our agent models the other agent's beliefs about the state of the world, and assumes that the other agent does not take into account our own beliefs, a common approach (Gmytrasiewicz and Doshi, 2005).

Agents make decisions according to a *policy* $\pi_i : \Delta(S) \rightarrow A$ which maximizes the discounted expected reward $\sum_{t=0}^{\infty} \gamma^t \mathbb{E}[R(s^t, a_1^t, a_2^t)|b_0, \pi_1, \pi_2]$. Using the assumption that the other agent tracks one less level of belief, we can solve for the other agent's policy $\bar{\pi}$, which allows us to estimate his actions and beliefs over time. To construct policies, we use Perseus (Spaan and Vlassis, 2005), a point-based value iteration algorithm.

Even tracking just one level of nested beliefs quickly leads to a combinatorial explosion in the number of belief states the other agent might have. This causes decision making in Dec-POMDPs to be NEXP complete, limiting their application to problems with only a handful of states (Bernstein et al., 2002). To ameliorate this difficulty, we use the method of Vogel et al. (2013), which creates a single-agent approximation to the full Dec-POMDP. To form this single-agent POMDP, we augment the state space to be $S \times S$, where the second set of state variables allows us to model the other agent's beliefs. We maintain a *point estimate* $\bar{b}$ of the other agent's beliefs, which is formed by summing out observations O that the other player might have received. To accomplish this, we factor the transition distribution into two terms: $T((s', \bar{s}')|a, \bar{\pi}(\bar{s}), (s, \bar{s})) = \bar{T}(\bar{s}'|s', a, \bar{\pi}(\bar{s}), (s, \bar{s}))T(s'|a, \bar{\pi}(\bar{s}), (s, \bar{s}))$. This observation marginalization can be folded into the transition distribution $\bar{T}(\bar{s}'|s', a, \bar{\pi}(\bar{s}), (s, \bar{s}))$:

$$\begin{aligned}\bar{T}(\bar{s}'|\ s', a, \bar{\pi}(\bar{s}), (s, \bar{s})) &= \Pr(\bar{s}'|s', a, \bar{\pi}(\bar{s}), (s, \bar{s})) \\ &= \sum_{\bar{o} \in O} \Bigg(\frac{\Omega(\bar{o}|\bar{s}', a, \bar{\pi}(\bar{s}))T(\bar{s}'|a, \bar{\pi}(\bar{s}), \bar{s})}{\sum_{\bar{s}''} \Omega(\bar{o}|\bar{s}'', a, \bar{\pi}(\bar{s}))T(\bar{s}''|a, \bar{\pi}(\bar{s}), \bar{s})} \\ &\quad \times \Omega(\bar{o}|s', a, \bar{\pi}(\bar{s})) \Bigg) \qquad (1)\end{aligned}$$

Communication is treated as another type of observation, with messages coming from a finite set M. Each message $m \in M$ has the semantics $\Pr(s|m)$, which represents the probability that the world is in state $s \in S$ given that m is true. Messages m received from a partner are combined with perceptual observations $o \in O$, to form a joint observation (m, o).

A *literal listener*, denoted L, interprets messages according to this semantics, without taking into account the beliefs of the speaker. L assumes that the perceptual observations and messages are conditionally independent given the state of the world. Using Bayes' rule, the literal listener's joint observation/message distribution is

$$\begin{aligned}\Pr((o, m)|s, s', a) &= \Omega(o|s', a) \Pr(m|s) \\ &= \Omega(o|s', a) \frac{\Pr(s|m) \Pr(m)}{\sum_{m' \in M} \Pr(s|m') \Pr(m')} \qquad (2)\end{aligned}$$

The $\Pr(m)$ prior over messages can be estimated from corpus data, but we use a uniform prior for simplicity.

A *literal speaker*, denoted S, produces messages according to the most descriptive term:

$$\pi_S(s) = \arg\max_{m \in M} p(s|m). \qquad (3)$$

The literal speaker does not model the beliefs of the listener.

To interpret implicatures, a *level-one listener*, denoted L(S), models the beliefs a literal speaker must have had to produce an utterance: $\Pr(m|s) = \mathbb{1}[\bar{\pi}_S(s) = m]$, where $\bar{\pi}_S$ is the level-one listener's estimate of the speaker's policy. In this setting, we denote the level-one listener's estimate of the speaker's belief as $\bar{s}$, yielding the belief update equation

$$\begin{aligned}&\Pr((o, m)|(s, \bar{s}), (s', \bar{s}'), a, \bar{\pi}_S(\bar{s})) = \\ &\quad \Omega(o|s', a)\mathbb{1}[\bar{\pi}_S(\bar{s}) = m] \qquad (4)\end{aligned}$$

The literal semantics of messages is not explicitly included in the level-one listener's belief update. Instead, when he solves for the literal speaker's policy $\bar{\pi}_S$, the meaning of a message is the set of beliefs that would lead the literal speaker to produce the utterance.

A *level-one speaker*, S(L), produces utterances to influence a literal listener, and a *level-two listener*, L(S(L)), uses two levels of belief nesting to interpret utterances as the beliefs that a level-one speaker might have to produce that utterance. At each level of nesting, we apply the marginalized

r_1 r_2 r_3

(a) Scenario.

Message	r_1	r_2	r_3
moustache	$\frac{1}{2}$	$\frac{1}{2}$	0
glasses	0	$\frac{1}{2}$	$\frac{1}{2}$
hat	0	0	1

(b) Literal interpretations.

Message	r_1	r_2	r_3
moustache	1	0	0
glasses	0	1	0
hat	0	0	1

(c) Implicature-rich interpretations.

Figure 1: A simple reference game. The matrices give distributions $\Pr(t = r_i|\text{utterance})$

belief-state approach of (Vogel et al., 2013), augmenting the state space with another copy of the underlying world state space, where the new copy represents the next level of belief. For instance, the L(S(L)) agent will make decisions in the $S \times S \times S$ space. For an L(S(L)) state $(s, \bar{s}, \hat{s})$, s is the true state of the world, $\bar{s}$ is the speaker's belief of the state of the world, and $\hat{s}$ is the speaker's belief of the listener's beliefs. In the next two sections we show how a level-one and level-two listener infer implicatures.

3 Reference Game Implicatures

Fig. 1a is the scenario for a reference game of the sort pioneered by Rosenberg and Cohen (1964) and Dale and Reiter (1995). The potential referents are r_1, r_2, and r_3. Speakers use a restricted vocabulary consisting of three messages: 'moustache', 'glasses', and 'hat'. The speaker is assigned a referent r_i (hidden from the listener) and produces a message on that basis. The speaker and listener share the goal of having the listener identify the speaker's intended referent r_i.

Fig. 1b depicts the literal interpretations for this game. It looks like the listener's chances of success are low. Only 'hat' refers unambiguously. However, the language and scenario facilitate *scalar implicature* (Horn, 1972; Harnish, 1979; Gazdar, 1979). Briefly, the scalar implicature pattern is that a speaker who is knowledgeable about the relevant domain will choose a communicatively weak utterance U over a communicatively stronger utterance U' iff U' is false (assuming U and U' are relevant). The required sense of communicative strength encompasses logical entailments as well as more particularized pragmatic partial orders (Hirschberg, 1985).

In our scenario, 'hat' is stronger than 'glasses': the referents wearing a hat are a proper subset of those wearing glasses. Thus, given the players' goal, if the speaker says 'glasses', the listener should draw the scalar implicature that 'hat' is false. Thus, 'glasses' comes to unambiguously refer to r_2 (Fig. 1c, line 2). Similarly, though 'moustache' and 'glasses' do not *literally* stand in the specific–general relationship needed for scalar implicature, they do with 'glasses' *pragmatically* associated with r_2 (Fig. 1c, line 1).

Our implementation of these games as Dec-POMDPs mirrors their intuitive description and their treatment in iterated best response models (Jäger, 2007; Jäger, 2012; Franke, 2009; Frank and Goodman, 2012). The state space S encodes the attributes of the referents (e.g., $\textbf{hat}(r_2) = \mathsf{T}$, $\textbf{glasses}(r_1) = \mathsf{F}$) and includes a target variable t identifying the speaker's referent (hidden from the listener). The speaker has three speech actions, identified with the three messages. The listener has four actions: 'listen' plus a 'choose' action c_i for each referent r_i. The set of observations O is just the set of messages (construed as utterances). The agents receive a positive reward iff the listener action c_i corresponds to the speaker's target t. Because this is a one-step reference game, the transition distribution T is the identity distribution.

The *literal listener* L interprets utterances as a truth-conditional speaker would produce them (Fig. 1b). The *level-one speaker* S(L) augments the state space with a variable 'listener_target' and models L's beliefs $\bar{b}$ using the approximate methods of Sec. 2. Crucially, the optimal speaker policy $\pi_{\text{S(L)}}$ is such that $\pi_{\text{S(L)}}(t{=}r_3) =$ 'hat' and $\pi_{\text{S(L)}}(t{=}r_1) =$ 'moustache'. The *level-two listener* L(S(L)) models S(L) via an estimate of the 'listener_target' variable. For each speech action m, L(S(L)) considers all values of t and the likeli-

hood that S(L) would have produced m:

$$\Pr(t{=}r_i|m) \propto \mathbb{1}[\bar{\pi}_{\text{S(L)}}(t{=}r_i) = m]$$

Since S(L) uses 'hat' to describe r_3 and 'moustache' to describe r_1, L(S(L)) correctly infers that 'glasses' refers to r_2, completing Fig. 1c's full implicature-rich pattern of mutual exclusivity (Clark, 1987; Frank et al., 2009).

This basic pattern is robustly attested empirically in human data. The experimental data are, of course, invariably less crisp than our idealized model predicts, but many important sources of variation could be brought into our model, with the addition of strong salience priors (Frank and Goodman, 2012; Stiller et al., 2011), assumptions about bounded rationality (Camerer et al., 2004; Franke, 2009), and a 'soft-max' view of the listener (Frank et al., 2009).

4 Cards World Implicatures

The Cards corpus[1] contains 1266 metadata-rich transcripts from a two-player chat-based game. The world is a simple maze in which a deck of cards has been distributed. The players' goal is to find specific subsets of the cards, subject to a variety of constraints on what they can see and do. The Dec-POMDP-based agents of Vogel et al. (2013) play a simplified version in which the goal is to be co-located with a single card. Vogel et al. show that their agents' linguistic behavior is broadly Gricean. However, their agents' language is too simple to reveal implicatures. The present section remedies this shortcoming. Implicature-rich interpretations are an immediate consequence.

We implement the simplified Cards tasks as follows. The state space S is composed of the location of each player and the location of the card. The transition distribution $T(s'|s, a_1, a_2)$ encodes the outcome of movement actions. Agents receive one of two sensor observations, indicating whether the card is at their current location. The players are rewarded when they are both located on the card. Each player begins knowing his own location, but not the location of the other player nor of the card.

The players have four movement actions ('up', 'down', 'left', 'right') and nine speech actions interpreted as identifying card locations. Fig. 2 depicts these utterances as a partial order determined by entailment. These general-to-specific relationships show that the language can support scalar conversational implicatures.[2]

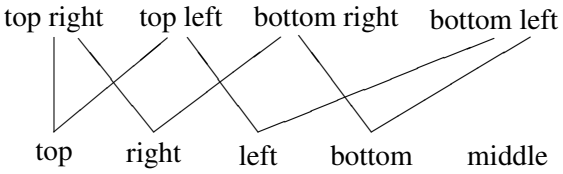

Figure 2: Cards world utterance actions.

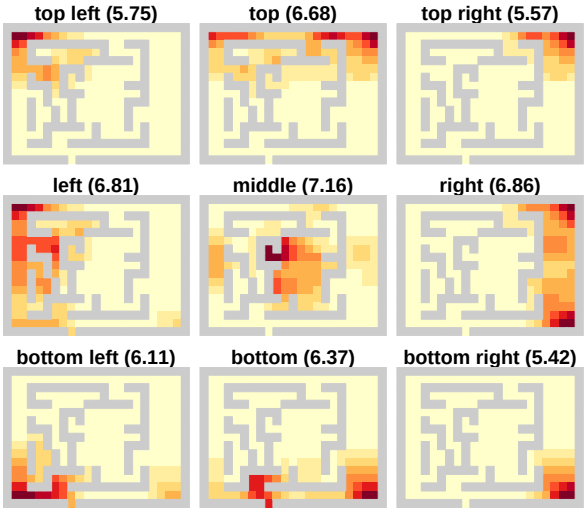

Figure 3: Literal interpretations derived from the Cards corpus. The entropy of each distribution is included in parentheses. Each term is estimated from all tokens that contain it, which washes out implicature-rich usage, thereby providing our model with an empirically-grounded literal start.

Fig. 2 is not entirely appropriate in our setting, however. Our expressions are vague; there is no sharp boundary between, e.g., 'top' and 'bottom', nor is it clear where 'top right' begins. To model this vagueness, we analyze each message m as denoting a conditional distribution $\Pr(x|m)$ over grid squares x in the gameboard. These distributions are derived from human–human Cards interactions using the data and methods of Potts (2012). Of course, there is a tension here: our model assumes that we begin with literal interpretations, but human–human data will reflect pragmatically-enriched usage. To get around this, we approximate literal interpretations by deriving each term's distribution from all the corpus tokens that contain it. For example, the distribution for 'top' is

[1] http://cardscorpus.christopherpotts.net

[2] Our agents cannot produce modified versions of 'middle' like 'middle right'. These would be synonymous with implicature-enriched general terms. We work with a simple cost-function that treats all forms alike, but future versions of this work will incorporate more realistic form-based costs.

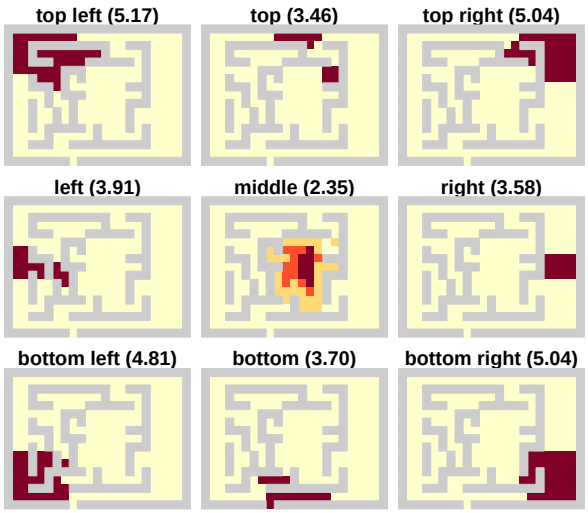

Figure 4: Implicature-rich interpretations, derived using the level-one listener L(S).

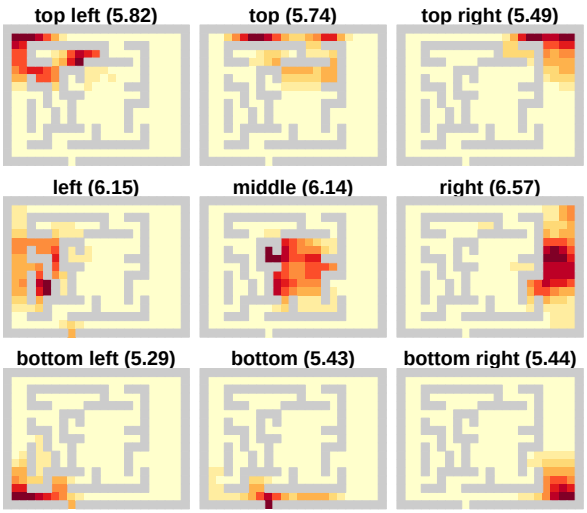

Figure 5: Distributions reflecting human speakers' aggregate referential intentions . Each term is estimated only from tokens that exactly match it.

estimated not only from 'top' but also from 'top right', 'middle right', and so forth. The denotation for 'top right' excludes simple 'top' and 'right' utterances but includes expressions like 'very top right'. This semantics washes out any implicature patterns, thereby giving us a proper literal starting point. Fig. 3 shows these denotations for the full set of expressions. The entailment relations from Fig. 2 are (fuzzily) evident. For example, the areas of high probability for 'right' properly contain the areas of high probability for 'top right'.

To show how the Dec-POMDP model delivers implicatures, we begin with a *literal speaker* S who does not consider the location of the other player and instead searches the board until he finds the card. After finding it, he communicates the referring expression with highest literal probability for his location, using the distributions from Fig. 3. We denote the literal speaker's policy by π_S. The *level-one listener* L(S) tracks an estimate of S's location and beliefs about the card location. Using the approximation defined in Sec. 2, L(S) interprets an utterance m as $\Pr(m|s) = \mathbb{1}[\bar{\pi}_S(s) = m]$. Thus, the meaning of each m is the set of beliefs that S might have to produce this utterance. Fig. 4 shows how L(S) interprets each message. The meaning of general terms like 'top' and 'right' now exclude their modified counterparts. This is evident in the lack of overlap between high-probability areas and in the lower entropy values.

Direct evaluation of this result against the corpus data is not possible, because the corpus does not encode interpretations. However, we expect listener interpretations to align with speaker intentions, and we can gain insight into (aggregate) speaker intentions using our method for grounding referential terms. Whereas the literal interpretation for message m is obtained from all the tokens that contain it (Fig. 3), the speaker's *intended interpretation* for m is obtained from all of the tokens that exactly match it. For instance, the meaning of 'top' now excludes tokens like 'top left'. Fig. 5 shows these denotations, which mirror the distributions predicted by our model (Fig. 4). Thus, the L(S) model correctly infers the pragmatic meaning of referring expressions as used by human speakers, albeit in an idealized manner.

5 Future Work

We showed that implicatures arise in cooperative contexts from nested belief models. Our listener-centric implicatures must be combined with rational speaker behavior (Vogel et al., 2013) to produce general dialog agents. The computational complexity of Dec-POMDPs is prohibitive, and our approximations can be problematic for deep belief nesting. Future work will explore sampling-based approaches to belief update and decision making (Doshi and Gmytrasiewicz, 2009) to overcome these problems. These steps will move us closer to a computationally effective, unified theory of pragmatic enrichment and decision making.

Acknowledgements This research was supported in part by ONR grants N00014-10-1-0109 and N00014-13-1-0287 and ARO grant W911NF-07-1-0216.

References

Leon Bergen, Noah D. Goodman, and Roger Levy. 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*.

Daniel S. Bernstein, Robert Givan, Neil Immerman, and Shlomo Zilberstein. 2002. The complexity of decentralized control of Markov decision processes. *Mathematics of Operations Research*, 27(4):819–840.

Colin F. Camerer, Teck-Hua Ho, and Juin-Kuan Chong. 2004. A cognitive hierarchy model of games. *The Quarterly Journal of Economics*, 119(3):861–898, August.

Herbert H. Clark and Deanna Wilkes-Gibbs. 1986. Referring as a collaborative process. *Cognition*, 22(1):1–39.

Eve V. Clark. 1987. The principle of contrast: A constraint on language acquisition. In Brian MacWhinney, editor, *Mechanisms of Language Acquisition*, pages 1–33. Erlbaum, Hillsdale, NJ.

Robert Dale and Ehud Reiter. 1995. Computational interpretations of the Gricean maxims in the generation of referring expressions. *Cognitive Science*, 19(2):233–263.

Judith Degen and Michael Franke. 2012. Optimal reasoning about referential expressions. In *Proceedings of SemDIAL 2012*, Paris, September.

David DeVault and Matthew Stone. 2007. Managing ambiguities across utterances in dialogue. In Ron Artstein and Laure Vieu, editors, *Proceedings of DECALOG 2007: Workshop on the Semantics and Pragmatics of Dialogue*.

Prashant Doshi and Piotr J. Gmytrasiewicz. 2009. Monte carlo sampling methods for approximating interactive pomdps. *J. Artif. Int. Res.*, 34(1):297–337, March.

Michael C. Frank and Noah D. Goodman. 2012. Predicting pragmatic reasoning in language games. *Science*, 336(6084):998.

Michael C. Frank, Noah D. Goodman, and Joshua B. Tenenbaum. 2009. Using speakers' referential intentions to model early cross-situational word learning. *Psychological Science*, 20(5):579–585.

Michael Franke. 2009. *Signal to Act: Game Theory in Pragmatics*. ILLC Dissertation Series. Institute for Logic, Language and Computation, University of Amsterdam.

Gerald Gazdar. 1979. *Pragmatics: Implicature, Presupposition and Logical Form*. Academic Press, New York.

Piotr J. Gmytrasiewicz and Prashant Doshi. 2005. A framework for sequential planning in multi-agent settings. *Journal of Artificial Intelligence Research*, 24:24–49.

Dave Golland, Percy Liang, and Dan Klein. 2010. A game-theoretic approach to generating spatial descriptions. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 410–419, Cambridge, MA, October. ACL.

H. Paul Grice. 1975. Logic and conversation. In Peter Cole and Jerry Morgan, editors, *Syntax and Semantics*, volume 3: Speech Acts, pages 43–58. Academic Press, New York.

Robert M. Harnish. 1979. Logical form and implicature. In *Linguistic Communication and Speech Acts*, pages 313–391. MIT Press, Cambridge, MA.

Julia Hirschberg. 1985. *A Theory of Scalar Implicature*. Ph.D. thesis, University of Pennsylvania.

Laurence R Horn. 1972. *On the Semantic Properties of Logical Operators in English*. Ph.D. thesis, UCLA, Los Angeles.

Gerhard Jäger. 2007. Game dynamics connects semantics and pragmatics. In Ahti-Veikko Pietarinen, editor, *Game Theory and Linguistic Meaning*, pages 89–102. Elsevier, Amsterdam.

Gerhard Jäger. 2012. Game theory in semantics and pragmatics. In Maienborn et al. (Maienborn et al., 2012).

Emiel Krahmer and Kees van Deemter. 2012. Computational generation of referring expressions: A survey. *Computational Linguistics*, 38(1):173–218.

Claudia Maienborn, Klaus von Heusinger, and Paul Portner, editors. 2012. *Semantics: An International Handbook of Natural Language Meaning*, volume 3. Mouton de Gruyter, Berlin.

Christopher Potts. 2012. Goal-driven answers in the Cards dialogue corpus. In Nathan Arnett and Ryan Bennett, editors, *Proceedings of the 30th West Coast Conference on Formal Linguistics*, Somerville, MA. Cascadilla Press.

Hannah Rohde, Scott Seyfarth, Brady Clark, Gerhard Jäger, and Stefan Kaufmann. 2012. Communicating with cost-based implicature: A game-theoretic approach to ambiguity. In *The 16th Workshop on the Semantics and Pragmatics of Dialogue*, Paris, September.

Seymour Rosenberg and Bertram D. Cohen. 1964. Speakers' and listeners' processes in a word communication task. *Science*, 145:1201–1203.

Matthijs T. J. Spaan and Nikos Vlassis. 2005. Perseus: Randomized point-based value iteration for POMDPs. *Journal of Artificial Intelligence Research*, 24(1):195–220, August.

Alex Stiller, Noah D. Goodman, and Michael C. Frank. 2011. Ad-hoc scalar implicature in adults and children. In *Proceedings of the 33rd Annual Meeting of the Cognitive Science Society*, Boston, July.

Adam Vogel, Max Bodoia, Dan Jurafsky, and Christopher Potts. 2013. Emergence of Gricean maxims from multi-agent decision theory. In *Human Language Technologies: The 2013 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, Atlanta, Georgia, June. Association for Computational Linguistics.