Abductive understanding of dialogues about joint activities

Pat Langley[1,2], Ben Meadows[2], Alfredo Gabaldon[1] &
Richard Heald[2]
[1]Carnegie Mellon University / [2]University of Auckland

This paper examines the task of understanding dialogues in terms of the mental states of the participating agents. We present a motivating example that clarifies the challenges this problem involves and then outline a theory of dialogue interpretation based on abductive inference of these unobserved beliefs and goals, incremental construction of explanations, and reliance on domain-independent knowledge. After this, we describe UMBRA, an implementation of the theory that embodies these assumptions. We report experiments with the system that demonstrate its ability to accurately infer the conversants' mental states even when some speech acts are unavailable. We conclude by reviewing related research on dialogue and discussing avenues for future study.

Keywords: dialogue understanding; abductive inference; mind-reading; mental state ascription

1. Introduction

The ability to participate in extended dialogues is one of the most distinctive characteristics of human intelligence. By letting people communicate their beliefs, goals, and intentions, such conversations support the coordination of complex joint activities, making it possible to achieve objectives that individuals cannot accomplish on their own. Yet the processes that underlie this capacity involve more than those for understanding and generating individual utterances. Because human dialogues leave so much unsaid, they also require representations and mechanisms for inferring the mental states of other participants from incomplete information. In other words, dialogue understanding is centrally about 'mind reading'.

In this paper, we explore these intriguing aspects of dialogue. We will not deal with speech processing, which is already a mature technology, or with sentence-level processing, which has also been studied extensively. Instead, we will focus

ISSN 1572–0373 / E-ISSN 1572–0381

on the mechanisms that occur at levels above these more basic operations. Others have given computational treatments of conversation at this level (e.g. Perrault & Allen 1980; Litman 1985), including some that address the process of drawing inferences about others' goals and beliefs; we are focusing our attention on the particular area of mind-reading in the context of collaborative dialogue. Advances here would fill a gap in our understanding of language and social interaction.

The computational system we describe here does not process spoken language, carry out syntactic analysis, or map from sentences to their logical meanings, since we provide the latter directly. Our system does not carry out dialogues at the logical level, but rather attempts to understand others' conversations. We argue that this is appropriate if our aim is to clarify the representations and processes of mind reading, which is the main focus of this paper. Ignoring the lower levels of dialogue lets us focus on this topic.

We will limit our attention to dialogues about joint activities in which the participants are working toward shared aims. The most remarkable aspect of such interactions is how little the conversants state explicitly and how much they read between the lines. We maintain that this ability depends on a rich representation of the other agents' mental states and an abductive reasoning mechanism that draws plausible inferences about those states. Moreover, we claim that, although effective dialogue undoubtedly depends on shared knowledge about the domain under discussion, many important inferences draw on dialogue-level knowledge that does not refer to any domain predicates, making it relevant to many different settings.

In the remaining pages, we report our progress toward understanding this distinctive human capability in computational terms. We start by presenting an illustrative dialogue that clarifies the issues we must address. After this, we present a high-level theory of dialogue understanding, along with its implementation in an abductive reasoning system that infers the mental states of participants from a sequence of their speech acts. Our theory incorporates many ideas from earlier work on the logical analysis of dialogue, but it brings them together in a unified account that constitutes an advance in its own right. We also present an evaluation that demonstrates the system's ability to work from incomplete information, which is an intrinsic feature of human dialogues. In closing, we discuss related work and outline directions for future research.

2. A motivating dialogue

We can illustrate the class of problems that we plan to address with a simple dialogue that involves two agents: a human medic in a battlefield setting and an expert who attempts to assist him, through a remote audio link, in treating an injured person. This scenario is plausible because battlefield medics undergo limited training and they often encounter situations in which they can benefit from expert advice. We assume that the medic and expert each know about and trust the other.

> Medic: We have a man injured!
> Expert: Where is he hurt?
> Medic: He's bleeding from the left leg.
> Expert: How bad is the bleeding?
> Medic: Pretty bad. I think it's the artery.
> Expert: Okay, use a tourniquet to stop the bleeding.
> Medic: Right, where shall I put it?
> Expert: Just below the joint above the wound.
> Keep turning until it stops bleeding.
> Medic: Okay, the bleeding has stopped.

Despite this dialogue's simplicity, it raises many issues about the participants' mental states, about how to represent them, and about how they change over time. We maintain that a computational system which can interpret this conversation will generalize to other dialogues that involve joint activities in which one agent is attempting to help another achieve shared goals.

Analyzing this dialogue informally will clarify the types of inferences that it involves. Superficially, the first utterance appears to simply provide information that an injury has occurred, but deeper inspection suggests it also contains an implicit proposal that the expert help treat the new patient. Similarly, the expert's response at first seems to involve only a question, but it also contains an implicit acknowledgment of the injury and an implicit agreement to help with the problem. There is also an intrinsic coherence to the dialogue, with later utterances relating directly to earlier ones, providing either confirmations or elaborations about previous content.

Another important aspect of the dialogue, relevant to the topic of mind reading, is that each participant makes informed guesses about the other agent's beliefs and goals. For instance, after hearing the first utterance, the expert most likely believes not only that a soldier is injured, but that the medic wants him to believe that and, after he has made the statement, believes that the expert believes it. Similarly, after the expert has asked the first question, the medic probably believes that the expert has adopted a goal to know the injury's location and that he expects the medic to have adopted the same goal.

3. Theoretical assumptions about dialogue understanding

In general, it is useful to distinguish theories from specific models that instantiate them. Theories comprise a set of abstract tenets that offer a framework within which more concrete models may be specified. The theory of dialogue understanding that we adopt in this paper incorporates four primary assumptions:

– *Dialogue understanding relies centrally on inference about the mental states of the participants*. These states include not only agents' beliefs and goals about the environment, but also their beliefs and goals about others' beliefs and goals.
– Because it involves the construction of plausible explanations, *dialogue understanding is inherently abductive in character*. Thus, it depends on postulating default assumptions about the conversing agents' beliefs and goals that, together, clarify relations among dialogue elements.[1]
– Because the utterances in a dialogue arrive sequentially, *the understanding mechanism operates in an incremental manner*. Thus, a few communicative acts are interpreted at a time, with later inferences in the explanation building upon ones introduced earlier.[2]
– Although dialogue understanding is a knowledge-driven process, *much of the responsible knowledge is meta-level in that it makes no reference to domain predicates*, so that the same structures support conversation across many different domains. Yet this dialogue-level knowledge can interact synergistically with domain-level content to improve the explanation process.

Taken together, these four assumptions provide an initial theory for the understanding of dialogues about joint activities. Although the theoretical postulates are not themselves directly testable, we can evaluate their viability by incorporating them into a running program that interprets dialogues from known speech acts. As we discuss later, this lets us propose claims that are subject to empirical tests, which in turn can offer indirect support for the framework.

Our theoretical views about dialogue have been influenced by a number of earlier treatments of this topic. For instance, Clark's (1996) naturalistic studies of conversation emphasize the incremental construction of *common ground* among participants, including beliefs about others' beliefs. We have also been inspired by Perrault and Allen's (1980) early logical analyses of dialogues, which describe speech acts in terms of their effects on others' beliefs and goals. Other work in this tradition – by Litman (1985), by Carberry and Lambert (1999), and

1. Some researchers consider *analogical reasoning* the core mechanism in human cognition, including dialogue understanding. Analogical reasoning is, in our view, a form of abductive reasoning.

2. This is not strictly required, as one could retain previous utterances and construct a new explanation whenever new ones arrive, but this seems unlikely to scale well for extended conversations and it would violate our intuitions about human dialogue understanding.

by McRoy and Hirst (1995) – has adopted similar assumptions about representations, knowledge, and processing for dialogue.

The approach we report in the remaining pages, although indebted to these early efforts, moves beyond them to embed their ideas in a general architecture for abductive inference that we have applied to other tasks, including understanding single-agent plans (Meadows, Langley, & Emery 2013a) and multi-agent stories (Meadows, Langley, & Emery 2013b). We will not focus on these tasks here, but our use of the same representations and mechanisms for dialogue suggests that our account is a general one that goes beyond its predecessors.

4. An abductive approach to dialogue interpretation

Our scientific aim is a computational account of dialogue understanding. In this section, we describe an implemented system, UMBRA, that incorporates the theoretical tenets just described. We start by describing the content of the system's working memory and knowledge base, after which we discuss the mechanisms that operate over them. We conclude with an example of the system's operation. We have implemented UMBRA in SWI-Prolog (Wielemaker et al. 2012) because it supports relational logic and embedded structures, and also because its ability to assert and retract facts (combined with SWI global variables) supports a working memory that changes over time. The default Prolog engine is limited to deductive proofs. Rather than using this native functionality, we implement a new layer (described in Section 4.3) on top of it so that the system can perform abductive inference.

4.1 Beliefs, goals, and speech acts

UMBRA distinguishes between domain-specific predicates, which describe content about particular environmental states and activities, and meta-level predicates, which describe an agent's views about such content. Predicates such as *has-injury* and *apply-tourniquet* are examples of the former, while the predicates *belief*, *goal*, and *constraint* are instances of the latter. Meta-level relations play a key role in describing the inferred mental states of participants in a dialogue.

Literals that involve meta-level predicates, as well as some that involve domain predicates, hold over a temporal interval in the sense that they become true at some time *T1* and false at some later time *T2*. Any of a literal's arguments may be unbound at any given point. Thus,

(1) belief(medic, has-injury(p1, i1), 09:00, 09:30)

means that the medic believes, from time 09:00 to 09:30, that the person p1 has an injury i1, while

(2) goal(expert, is-stable(p1), 09:05, T)

denotes that the expert adopted the goal at time 09:05 for p1 to be stable, and holds that goal until some unknown future time T. Similarly,

(3) constraint(expert, after(X, 09:05), 09:10, Y)

constrains the expert, from time 09:10, to regard the time X as occurring after time 09:05. We consider constraints to be element-level mental phenomena, as they can refer equally to the content of beliefs, to goals, or to other constraints.

Each of these mental state descriptions may itself be embedded, as in the statement

(4) belief(medic, goal(expert, is-stable(p1, T), 09:05, T), 09:15, 09:30),

which means that the medic believes from 09:15 that the expert has a goal from 09:05 to T for p1 to be stable. We maintain that a notation which lets one embed temporally-bound mental state elements within each other, combined with a formalism for communicative actions and a concomitant set of domain knowledge, provides the basic machinery needed to represent dialogues about joint activities.

UMBRA has a *working memory* consisting of a set of elements of this type, stored as Prolog literals that describe aspects of the world. All mental states are associated with some agent; elements at the top level of working memory correspond to those of the primary agent, i.e. the system itself. Working memory elements may describe an agent's beliefs or goals about the environment, as in (1) to (3) above, or they may encode its beliefs, goals, or constraints about another agent's beliefs or goals, as in (4). Domain-level predicates, such as *has-injury*(*p1, i1*) and *apply-tourniquet*(*tourniquet1, left-leg*), provide crucial content but do not appear outside the beliefs and goals of agents who are participating in the dialogue or beliefs of the primary agent engaged in the explanation task.

Analyses of dialogues often revolve around the notion of 'speech acts' (Austin 1962; Searle 1969), that is, conversational steps that produce certain mental effects in the participants. UMBRA may have elements in working memory that involve beliefs and goals about such activities. There are many distinct taxonomies of speech acts, and we will not take a position here about their relative merits. Instead, we have limited our model to six basic types of speech acts that are reasonably uncontroversial and that appear to be sufficient building blocks for the system to handle dialogues of various levels of nuance. These speech acts are:

- *Inform*, in which the speaker conveys some content to the listener;
- *Acknowledge*, in which the speaker states that he accepts some content;
- *Question*, in which the speaker asks the listener to provide some content;
- *Propose*, in which the speaker asks the listener to adopt some goal;

- *Accept*, in which the speaker tells the listener he has adopted a goal; and
- *Reject*, in which the speaker tells the listener he has declined to adopt a goal.

There are also auxiliary speech acts such as those that denote the beginning or end of a dialogue, or which express that an agent does not know the answer to a question.

Our system adopts a formal notation that, for each conversational step, specifies the type, the speaker, the listener, the content being communicated, and the time at which it occurred. For example, the literal

inform(medic, expert, has-injury(p1, i1), 10:15, 10:16)

denotes that the medic has informed the expert that "p1 has an injury i1", where the third argument indicates the content being communicated, and this took place between 10:15 and 10:16. Similarly, the speech act instance

question(expert, medic, location(i1), 05:33, 05:34)

denotes that from 05:33 to 05:34, the expert asked the medic a question about the location of the injury. As with domain-level literals, UMBRA always embeds speech acts inside *belief* or *goal* predicates. For instance, *belief*(*medic, inform*(*medic, expert, has-injury*(*p1, i1*)*, 10:15, 10:16*)*, 10:16, T*), states that the medic believes from 10:16 that he has just informed the expert about an injury. We typically draw from the set of beliefs of this sort to provide UMBRA with 'observations' about speech acts in the dialogue.

We make a fundamental distinction between an utterance made by a participant and the speech acts that he or she intends to convey in that utterance. For example, the medic's exclamation "We have a man injured!" can be seen as combining an *Inform* act about the injury with a *Propose* act about the goal of stabilizing the patient. Our account does not attempt to explain how an agent infers particular speech acts from utterances; this is an important problem that deserves attention, but it is not one of our immediate objectives. We will also assume that the listener correctly interprets the speech acts intended by the speaker, so that no communication errors take place. Finally, because we are interested in cooperative activities, we assume the dialogue involves no deception, although this is an avenue we are pursuing in recent work, and (for example) Bridewell and Isaac (2011) have proposed a representational framework that handles deceptive conversations in terms of agents' models of others' mental states.

4.2 Dialogue-level and domain-level knowledge

As we have mentioned, our theory assumes that dialogue understanding depends on two forms of knowledge. One type involves domain-specific rules that support inference about goal-directed activities. For the medic scenario, this would include rules about injuries, bleeding, body parts, tourniquets, medical procedures, and the like. A more interesting form of knowledge involves domain-independent rules about speech acts. UMBRA operates over rules that associate each type of speech act with a distinct pattern of belief and goals. In most cases, the structure of these patterns is independent of the content they communicate, in that they do not refer to any domain predicates, and thus constitute meta-level knowledge.

Table 1 presents simplified versions of rules for two of the six types of speech acts, Inform and Reject, in our current implementation. In the table, the first line of each rule is its head; the remainder is a set of antecedents that describe a relational pattern that is associated with this head. This pattern comprises beliefs and goals held by the speaker and listener, along with a primitive communicative action such as *inform-utterance* and *reject-utterance*. For example, the first rule in Table 1 encodes an Inform act (taking place from time T1 to T2) in which a Speaker informs a Listener of some Content. The pattern of concepts associated with this speech act involves a goal of the Speaker for the Listener to believe the Content, the actual fact of the Speaker making the utterance, the Speaker's ensuing belief that the Listener comes to believe the Content, and so on. The second rule encodes the Reject act analogously.

Table 1. The simplified forms of dialogue-level rules encoding two of the speech acts, *Inform* and *Reject*. Variables are written with initial capitals. Expressions of the form (*T*<*T'*) denote constraints on temporal variables.

inform(Speaker, Listener, Content, T1, T2) ←
  goal(Speaker, belief(Listener, Content, T2, T3), T4, T5,
  inform-utterance(Speaker, Listener, Content, T1, T2),
  belief(Speaker, belief(Listener, Content, T2, T6), T2, T7),
  belief(Listener, goal(Speaker, belief(Listener, Content, T2, T8), T9, T10), T2, T11),
  belief(Listener, belief(Speaker, Content, T12, T13), T2, T14),
  belief(Listener, Content, T2, T15),
  (*Speaker ≠ Listener*), (*T4 < T1*), (*T2 < T5*), (*T9 < T1*), (*T2 < T10*), (*T12 < T1*), (*T2 < T13*).

reject(Speaker, Listener, Content, T1, T2) ←
  not(goal(Speaker, Content, T3, T4)),
  goal(Speaker, belief(Listener, not(goal(Speaker, Content, T5, T6)), T2, T7), T8, T9),
  reject-utterance(Speaker, Listener, Content, T1, T2),
  belief(Speaker, goal(Listener, goal(Speaker, Content, T10, T11), T12, T2), T2, T13),
  belief(Speaker, belief(Listener, not(goal(Speaker, Content, T14, T15)), T2, T16), T2, T17),
  belief(Listener, not(goal(Speaker, Content, T18, T19)), T2, T20),
  (*Speaker ≠ Listener*), (*T3 < T1*), (*T2 < T4*), (*T5 < T1*), (*T2 < T6*), (*T8 < T1*), (*T2 < T9*),
  (*T10 < T1*), (*T12 < T1*), (*T14 < T1*), (*T2 < T15*), (*T18 < T1*), (*T2 < T19*).

Each rule also includes constraints, most of which specify orderings on times associated with antecedents. For simplicity, we have omitted implicit constraints that state that the start time on each is prior to its end time. Many of a rule's time constraints distinguish what are traditionally viewed as conditions and effects. Conditions have an end time that corresponds to the rule's start, effects have a start time that corresponds to the rule's end, and invariants have start and end times that lie outside its start and end times. For example, *goal*(*Listener,* […], *T12, T2*) is a condition, *belief*(*Listener, Content, T2, T3*) is an effect, and *belief*(*Speaker, Content, T12, T13*) is an invariant.

Note that neither the Inform nor the Reject rule assumes processing from the particular perspective of a speaker, listener, or a third party, even though the beliefs, goals, and constraints of an agent *X* are associated with a conversational act performed by an agent *Y*. As we will see later, UMBRA applies rules from particular perspectives by automatically embedding their elements within appropriate mental structures. Thus, instances of speech-act rules will have their heads and conditions embedded in other mental states. The same holds for other rules, of both the meta-level and domain-level variety.

These rule structures contain important similarities and recurring patterns. For instance, one can transform the Inform rule into the Propose rule by switching certain belief and goal predicates; the Accept and Acknowledge rules have an analogous relationship; and the Reject rule is very similar to the Accept rule but with key elements negated. These rules place strong constraints on the relations between the speech acts in a dialogue and the mental states that one can assign to its participants.

Of course, there are important aspects of dialogue that speech acts in isolation do not address. For example, we know that a speaker only makes an acknowledgment in response to an inform statement, that a question should be followed by an answer, and that a proposal should be answered by an acceptance or rejection. Such patterns constitute a form of domain-independent knowledge about the structure of conversations, above the level of speech acts themselves, that constrains and guides the understanding process. We encode this knowledge in a *dialogue grammar* consisting of 15 high-level rules that specify a dialogue in terms of its constituent speech acts. This knowledge is similar in format to that for speech acts. Table 2 presents an abstract version of the grammar that omits the rules' time constraints. There are several types of primary constituent:

– The highest-level structures are *dialogue components*. Most of these dialogue constituents comprise ordered sequences or *exchanges* of related speech acts.[3]

3. The literature sometimes refers to these as *adjacency pairs* (Sacks et al. 1978), but our system encodes more than pairwise dependencies.

- In a *Propose-Response* exchange, one agent proposes some course of action, after which the other agent responds; in contrast, in an *Inform-Acknowledge* exchange, one agent makes a statement and the other acknowledges it. In either case, the exchange may include a third component in the form of a subdialogue between the speech acts.
- In a *Question-Answer* exchange, one agent requests new, relevant information, often a clarification about some content just mentioned. After this, the other agent provides an appropriate response, either an Inform or a declaration that an answer is unavailable. The exchange ends when the original agent Acknowledges the answer.
- A *Response* simply encodes either an Accept act or a Reject act in a generic way, so that the response may be used in higher level grammar structures without having to refer to the different predicates.
- A *Reject-Reason* exchange combines a Propose-Response exchange that involves a rejection followed by an Inform-Acknowledge exchange that contains a reason for this rejection in terms of relevant domain knowledge.

Table 2. Summary of the dialogue grammar rules provided to UMBRA. Here we provide the basic structure of the system's 15 higher-level rules (an additional nine speech act rules and 15 domain-level rules are not shown). The actual rules refer to the communicating agents, time stamps, and constraints. Variables expressed using the symbol *C* refer to a rule's domain-level content.

dialogue-component ← dialogue-open
dialogue-component ← question-answer-exchange(C_1, C_2)
dialogue-component ← reject-reason-exchange(C, Reason)
dialogue-component ← propose-response-exchange(C, Response)
dialogue-component ← inform-ack.-exchange(C)
dialogue-component ← dialogue-close

question-answer-exchange(C_1, C_2) ← question(C_1), inform-ack.-exchange(C_2), relevant(C_1, C_2)
question-answer-exchange(C, *unknown*) ← question(C), unknown-response(C), acknowledge(unknown-response)

reject-reason-exchange(C_1, C_2) ← propose-response-exchange(C_1, reject), inform-ack.-exchange(C_2), relevant(C_1, C_2)

propose-response-exchange(C, *Response*) ← propose(C), response(C, Response)
propose-response-exchange(C_1, *Response*) ← propose(C_1), question-answer-exchange(C_2, C_1), response(C_1, Response), relevant(C_1, C_2)

response(C, *accept*) ← accept(C)
response(C, *reject*) ← reject(C)

inform-ack.-exchange(C) ← inform(C), acknowledge(C)
inform-ack.-exchange(C_1) ← inform(C_1), question-answer-exchange(C_2, _), acknowledge(C_1), relevant(C_1, C_2)

Note that some of the constituents in the grammar rules, such as *inform*(C) and *acknowledge*(C), share the same argument, indicating that they refer to the same content. Other components, such as *inform*(C_1) and *question-subdialogue*(C_2), take distinct arguments but link them through a requirement that the content of the two literals be relevant. Predicates for relevance are, necessarily, defined at the domain level. For instance, the fact that a tourniquet has been used is relevant to its current location, and the extent of bleeding is relevant to the size of an injury. UMBRA's grammar rules thus constrain the combination of speech acts so that the explanation process relates, for example, answers to questions in a coherent manner.

Domain-level knowledge is also necessary to encode relations among more concrete predicates. Knowledge used in the medic dialogue includes *conceptual knowledge* – rules that encode the relationships between conceptual predicates at a less abstract level than dialogue. For example,

If a leg injury is bleeding badly, then the injury is arterial.

UMBRA also incorporates rules that are responsible for generating goals, such as

If you believe someone is injured, then your goal is to make them stable.

The system specifies these knowledge elements as rules in which some of the domain content may be embedded in belief, goal, and constraint predicates. Although they are written in the same logical form as the dialogue grammar, these rules differ considerably from the grammar rules, which relate the beliefs and goals of different agents in a consistent mental embedding.

Domain rules can interact with those higher-level rules through the contents of working memory. For example, the same belief about a bleeding injury that unifies with a condition in a conceptual rule that defines arterial bleeds may also unify with a condition in a dialogue rule that describes an Inform speech act.

Together, these knowledge components describe what it means for a series of utterances to constitute a *well-formed dialogue*. For example, if the logical representations of utterances and mental states which support *inform*(*Speaker, Listener, Content, T1, T2*), *acknowledge*(*Listener, Speaker, Content, T2, T3*) are present in working memory, then *inform-acknowledge-exchange*(*Speaker, Listener, Content, T1, T3*) is a valid structure to infer (a further simplified version of this appears in Table 2). If the utterances were preceded and followed by elements describing speech acts for *dialogue-open* and *dialogue-close* respectively, then the result is an ordered sequence of dialogue components, which constitutes a dialogue.

4.3 Abductive inference about dialogues

Now that we have described UMBRA's internal representations and the content they encode, we can turn to the mechanisms it uses to process them. The system's inputs take the form of new beliefs about the occurrences of particular utterances, in logical form, along with domain-level, speech-level and meta-level knowledge in hierarchical form. The system operates incrementally, in that it receives beliefs about speech acts on successive 'input cycles', adding its inferences to working memory so they are available to influence later reasoning.

The key outputs are UMBRA's inferences about the mental phenomena – beliefs, goals, and constraints about the agents and environment – that accompany each speech act. These may include beliefs about other speech acts that have not been observed (see Subsection 5.2 on elided utterances), and higher-level structures that are instances of the components in the dialogue grammar. Some assumed literals may contain variables that were not bound when a rule was applied – for example, UMBRA might assume *belief*(*medic, has-injury*(*p1, i1*), *09:00, _Var1*) without inferring any particular value for *_Var1*, the end time on the medic's belief. Together, these outputs form an *explanation* of observed events that takes the form of a directed graph of inferences. The nodes in this graph are elements in UMBRA's working memory, and the edges that link them denote elements' membership as conditions or heads in particular rule instances.

We will only give a précis of UMBRA's mechanisms here, as more thorough descriptions are available elsewhere (Meadows, Langley, & Emery 2013a, 2013b). The architecture builds an explanation incrementally as new inputs become available, chaining off input utterances (in logical form) and previously inferred elements in a data-driven fashion. This strategy creates the explanation from the bottom up, as the system attempts to infer which speech acts are taking place and show that they constitute a well-formed dialogue in terms of the hierarchical grammar rules. Inference is abductive in nature; UMBRA may introduce default assumptions in order to apply a rule whose conditions cannot all be matched with elements in memory. For example, abductive reasoning may generate beliefs about missing speech acts, such as tacit responses that are left unspoken.

Input cycles each involve one or more *inference cycles*.[4] On each inference cycle, UMBRA first identifies the set of rules with conditions or heads that unify with at least one element in working memory, and provisionally performs those unifications to generate a set of embedded rules with partially instantiated heads. It then expands these candidates by unifying any remaining rule antecedents with elements from working memory. When no suitable match is possible, due to a lack

4. These are somewhat analogous to recognize-act cycles in production system architectures (Neches, Langley, & Klahr 1987), although UMBRA's explanations are constructed incrementally through extension of existing inferences.

of appropriate elements or a contradiction with constraints, the system instead generates a default assumption that the rule condition holds.

UMBRA assigns each candidate a numeric cost according to an evaluation function which prefers rule instances that use fewer assumptions, both in absolute terms and as a fraction of the rule's antecedents. It also favors those incorporating a greater number of working memory elements that were previously not involved in any rule instances in the explanation.[5] For example, given rule instances with either four or five conditions, both requiring two assumptions, the system would prefer the second one. However, if the first rule used two elements that had not yet been incorporated into the explanation, it might prefer the first one. And if a third rule could be applied without making any assumptions, the evaluation function would also rank it highly.

The system selects the lowest-cost candidate for application, extending the explanation by adding the instantiated rule's elements to working memory, and reduces an abstract user-specified threshold by the selected rule's cost. If all candidates' costs exceed this threshold, the inference cycle fails – in which case the rule application is not performed and the overall input cycle ends, with a new one beginning when new inputs are given. Alternatively, if the inference cycle succeeds, the system enters a new inference cycle, using the lower threshold.[6]

This sequence of operations incrementally extends the explanation to incorporate ever more inputs, and, where necessary, adds assumptions as connective tissue. These additions, including the inferred rule heads, are immediately available for use in further inferences and can lead the system to replace variables in existing elements with constants, subject to basic checks for contradiction prevention. The end result is a coherent, connected, hierarchical account of the input utterances in terms of available background knowledge.

Note that the system expands explanations by monotonically extending working memory as it processes input: it never retracts elements it has added to memory.[7] However, the abductive inference mechanism makes default assumptions and hence is not guaranteed to make the correct inferences. This means it should be

5. The metric incorporates principles such as parsimony, consilience (Thagard 1978), and coherence (Ng & Mooney 1990).

6. Because the evaluation function assigns a nonzero cost to every candidate, a given input cycle is guaranteed to end regardless of the original threshold.

7. Some inference mechanisms involve 'non-monotonic reasoning', but this term is unrelated to our use of 'monotonic extension'. Because of its abductive character, our system does use non-monotonic reasoning in its technical sense: it is possible that, given a set of input observations $I1$, UMBRA will assume a proposition P, while given a set $I2, I1 \subset I2$, the system will not assume P.

possible to improve performance by adding processes for revising the system's beliefs when anomalies are detected, as we discuss in Section 7.

Table 3. UMBRA's additions to working memory when presented, at time 09:01, with literals encoding "We have a man injured!" from the medic dialogue. The prefix '_' denotes uninstantiated variables, in this case unknown time stamps. Italicized elements also appear in an instantiation of the Acknowledge rule in the explanation. Representation of the literals is simplified for clarity.

belief(UMBRA, inform-utterance(medic, expert, has-injury(p1, i1), 08:59, 09:00), 09:01, _1)
belief(UMBRA, goal(medic, belief(expert, has-injury(p1, i1), 09:00, _2), _3, _4), 09:01, _5)
belief(UMBRA, belief(medic, belief(expert, has-injury(p1, i1), 09:00, _6), 09:00, _7), 09:01, _8)
belief(UMBRA, belief(expert, goal(medic, belief(expert, has-injury(p1, i1), 09:00, _9), _10, _11), 09:00, _12), 09:01, _13)
belief(UMBRA, belief(expert, belief(medic, has-injury(p1, i1), _14, _15), 09:00, _16), 09:01, _17)
belief(UMBRA, belief(expert, has-injury(p1, i1), 09:00, _18), 09:01, _19)
constraint(UMBRA, (medic ≠ expert), 09:01, _20)
constraint(UMBRA, (_3 < 08:59), 09:01, _21)
constraint(UMBRA, (09:00 < _4), 09:01, _22)
belief(UMBRA, constraint(expert, (_10 < 08:59), 09:00, _23), 09:01, _24)
belief(UMBRA, constraint(expert, (09:00 < _11), 09:00, _25), 09:01, _26)
belief(UMBRA, constraint(expert, (_14 < 08:59), 09:00, _27), 09:01, _28)
belief(UMBRA, constraint(expert, (09:00 < _15), 09:00, _29), 09:01, _30)
belief(UMBRA, inform(medic, expert, has-injury(p1, i1), 08:59, 09:00), 09:01, _1)

Together, the mechanisms we have described implement the four theoretical tenets we stated in Section 3. Note that UMBRA does not need to construct an explanation with a single top-level rule instance, which distinguishes its accounts from those typically produced in work on plan recognition. The data-driven inference process may generate explanations that include multiple root nodes or none. As a result, the system works as well with dialogue fragments as it does with complete conversations.

We can clarify UMBRA's operation by examining its behavior on the initial portion of the sample medic dialogue we gave earlier. Table 3 presents one of the first inferences the system makes for this scenario. Recall the first utterance by the medic, "We have a man injured!" The system receives as input the logical form of that utterance with suitable time stamps:

inform-utterance(medic, expert, has-injury(p1, i1), 08:59, 09:00)

This observation is added to working memory in the form of a belief, shown in bold in Table 3. As the cycle continues, the system attempts to expand the explanation and to account for this new belief. After applying the embedding *belief*(*UMBRA*, · · ·) to the inform speech act rule (shown in Table 1) and instantiating *Speaker* with *medic*, *Listener* with *expert*, and *T1*, *T2* with the corresponding values from the new belief, UMBRA determines that the new belief unifies with the second element in the body of the instantiated rule. Because none of the other elements in the body match any elements in working memory, the system adds them to a set of assumptions and then checks whether the cost exceeds the threshold. It does not, so the system tentatively adds the assumptions to working memory and the inference cycle ends successfully. The remaining lines in Table 3 show the elements that are added to working memory as this cycle completes.

In the next inference cycle, UMBRA again looks for rule candidates to apply. Since the embedded head of the Inform act (the last line in Table 3) is in working memory, the system determines that it can apply the rule for an inform-acknowledge exchange,

inform-ack-exchange(Speaker, Listener, C)←
inform(Speaker, Listener, C), acknowledge(Listener, Speaker, C).

This incurs the small cost of making only two assumptions (the second body element and the head),[8] so the corresponding rule instance is applied to expand the explanation. UMBRA adds two elements to working memory,

belief(UMBRA, acknowledge(expert, medic, has-injury(p1, i1), 09:00, _32), 09:01, _33),
belief(UMBRA, inform-ack-exchange(medic, expert, has-injury(p1, i1), 08:59, _32), 09:01, _34),

plus constraints similar to those in Table 3. As the threshold has not been reached, the input cycle continues and more inferences can be made. One of the candidate rules the system considers next is the acknowledge speech act rule:

acknowledge(Speaker, Listener, Content, T1, T2) ←
belief(Speaker, Content, T3, T4)
acknowledge-utterance(Speaker, Listener, Content, T1, T2),
goal(Speaker, belief(Listener, belief(Speaker, Content, T5, T6), T2, T7), T8, T2),
belief(Speaker, belief(Listener, belief(Speaker, Content, T9, T10), T2, T11), T2, T12),
belief(Listener, belief(Speaker, Content, T13, T14), T2, T15),
(Speaker ≠ Listener),
(T3 < T1), (T2 < T4), (T5 < T1), (T2 < T6), (T9 < T1), (T2 < T10),
(T13 < T1), (T2 < T14).

8. For simplicity, time stamps and associated constraints are not shown here.

The first and the last (before the constraints) body elements of this rule match the working memory elements shown in italics in Table 3, with *Speaker* instantiated as *expert* and *Listener* as *medic*. The head of the Acknowledge rule also matches an element added when the rule for the inform-acknowledge exchange was applied. Thus, three elements are matched and three must be assumed, as constraints are not included in cost calculations. After checking that the constraints are satisfied and the cost is under the new threshold, the system adds the assumed elements to working memory. These elements, with some uninstantiated variables shown only as underscores, are:

> belief(UMBRA, acknowledge-utterance(expert, medic, has-injury(p1, i1), 09:00, _32), 09:01, _),
> belief(UMBRA, goal(expert, belief(medic, belief(expert, has-injury(p1, i1), _, _), _32, _), _32, _),
> belief(UMBRA, belief(expert, belief(medic, belief(expert, has-injury(p1, i1), _, _), _32, _), _32, _).

Note that one of the assumptions incorporated into the explanation is an acknowledge utterance that was elided from the actual dialogue and, as a result, from the sequence of inputs received (see Section 5.2 for further discussion of elided elements). From this point, the inference cycle proceeds until the threshold is reached.

As an aside, UMBRA also supports the generation of explanations from different agents' perspectives, inferring not what the primary agent believes the dialogue structure to be, but what the conversing agents (or an eavesdropper) would believe. This variant on mindreading is not particularly interesting for cases in which communication produces common ground, but is essential for computational treatments of deceptive dialogues (Bridewell & Isaac 2011) and identification of conversational misunderstandings (McRoy & Hirst 1995; Cahn & Brennan 1999).

In summary, UMBRA is presented with beliefs about the sequential occurrences of speech acts and, in response, it generates inferences about the participating agents' beliefs, goals, and constraints. To this end, the system uses an incremental form of abduction to construct an explanation that incorporates elements of the conversants' mental states as terminal nodes, many of which are default assumptions introduced to produce a coherent story. Together, these elements account for the higher levels of dialogue understanding and the role that mind reading plays therein.

5. Empirical results on dialogue understanding

Our approach to dialogue understanding seems promising, but we should also examine the extent to which UMBRA works as desired for this task. We wish to establish that our architecture has two main capabilities:

- When provided with sets of (the logical forms of) dialogue utterances, it generates reasonable inferences about those dialogues, including the mental states of the participating agents; and
- UMBRA can explain a dialogue even when some of the utterances are elided from the input – when some parts of the conversation are unspoken, implicit, or go unheard.

It is important to demonstrate that the framework exhibits these abilities, which appear central to dialogue interpretation.

Although the first capability may appear trivial, in fact we should not assume that the system will work perfectly when given all the relevant utterances. Consider first that our commitment to incremental processing means that, even when the full set of inputs is given, each observation only becomes available at a particular time. This means that, before the final input cycle, the system has access to only a portion of the inputs. Note also that rules share many conditions. Because of this, abductive inference may result in spurious rule applications when only partial input is available. Given that UMBRA is unaware whether it will be provided with all the possible inputs, it may well make mistakes when creating an account for the elements available at any given step in the process. The second ability – explaining elided dialogue – is even more important. In practice, many dialogues will feature elisions, as the ability to omit inferable elements makes human communication efficient.

To examine these two core capabilities, we developed four variants of the medic dialogue that we presented earlier. These variations are shown in Table 4. Scenario 1 is a simplified form in which no clarification about the tourniquet's location is required. Scenario 2 is amended to include a clarifying subdialogue between the initial proposal to use a tourniquet and its acceptance. Scenario 3 follows similar lines, but the medic rejects the expert's proposal to apply a tourniquet, stating that he has none to apply, so the expert proposes that the medic apply pressure above the wound manually. In Scenario 4, the medic cannot answer one of the expert's questions. The medic also rejects the expert's proposal, leading to a subdialogue in which the expert asks why the medic has no tourniquet and the medic clarifies that he has used it on another patient.

In order to demonstrate our system's abilities, we established the success criteria for the four scenarios in terms of which beliefs, goals, and constraints we

Table 4. The four dialogue scenarios. The utterances of the medic (M) and expert (E) are given for each variation. The system's input sequences were drawn from these scenarios in their logical forms.

Scenario 1 M: We have a man injured! E: Where is he hurt? M: He's bleeding from the left leg. E: How bad is the bleeding? M: Pretty bad. I think it's the artery. E: Okay, use a tourniquet to stop the bleeding. M: Right. E: Keep turning until it stops bleeding. M: Okay, the bleeding has stopped.	**Scenario 3** M: We have a man injured! E: Where is he hurt? M: He's bleeding from the left leg. E: How bad is the bleeding? M: Pretty bad. I think it's the artery. E: Okay, use a tourniquet to stop the bleeding. M: I can't – I don't have one. E: Try applying pressure manually instead. M: Okay, the bleeding has stopped.
Scenario 2 M: We have a man injured! E: Where is he hurt? M: He's bleeding from the left leg. E: How bad is the bleeding? M: Pretty bad. I think it's the artery. E: Okay, use a tourniquet to stop the bleeding. M: Right. Where should I put it? E: Below the knee. M: Okay. E: Keep turning until it stops bleeding. M: Okay, the bleeding has stopped.	**Scenario 4** M: We have a man injured! E: Where is he hurt? M: He's bleeding from the left leg. E: Has the artery been hit? M: I don't know. E: Okay. Use a tourniquet to stop the bleeding. M: I can't – I don't have one. E: What happened to it? M: I used it on another patient already. E: Okay. Try applying pressure manually instead. M: Okay, the bleeding has stopped.

believed the participants should adopt in the course of the conversation. Each scenario involves the inference of approximately 400 elements, including 30 to 40 instantiated speech acts and dialogue grammar rules, such as *propose*(*expert, medic, turn-tourniquet*(*tq1*)*, 09:10, 09:11*) and *propose-response-exchange*(*turn-tourniquet*(*tq1*)*, accept, 09:10, 09:12*). We then ran UMBRA on the sequence of input utterances from each scenario. As an example, Table 5 gives the full trace of input observations for Scenario 1. We recorded the inferences the system made and compared these to the target inferences. From this we found the number of correctly inferred elements (true positives), errors of commission (false positives), and errors of omission (false negatives). These in turn let us compute *precision* (informally, the proportion of the system's inferences that were good) and *recall* (the proportion of the good inferences that were made). Table 6 presents the results of these tests, with rows corresponding to scenarios and columns to scores.

Table 5. Set of input observations for Scenario 1 from Table 4. Time stamps, background beliefs, and the outermost predicates denoting beliefs of the system have been removed for readability.

```
belief(medic, initiate-dialogue(medic, expert, establish-radio-contact))
belief(medic, inform-utterance(medic, expert, is-injured(p1, i1)))
belief(medic, acknowledge-utterance(expert, medic, is-injured(p1, i1)))
belief(medic, propose-utterance(medic, expert, stable(p1)))
belief(medic, accept-utterance(expert, medic, stable(p1)))
belief(medic, question-utterance(expert, medic, location(i1, lc1)))
belief(medic, inform-utterance(medic, expert, where(lc1, left-leg)))
belief(medic, acknowledge-utterance(expert, medic, where(lc1, left-leg)))
belief(medic, inform-utterance(medic, expert, bleeding(i1, b1)))
belief(medic, acknowledge-utterance(expert, medic, bleeding(i1, b1)))
belief(medic, question-utterance(expert, medic, extent-of-bleed(b1, extent1)))
belief(medic, inform-utterance(medic, expert, size(extent1, large)))
belief(medic, acknowledge-utterance(expert, medic, size(extent1, large)))
belief(medic, inform-utterance(medic, expert, location(b1, artery1)))
belief(medic, acknowledge-utterance(expert, medic, location(b1, artery1)))
belief(medic, propose-utterance(expert, medic, apply-tourniquet(tq1, left-leg)))
belief(medic, accept-utterance(medic, expert, apply-tourniquet(tq1, left-leg)))
belief(medic, propose-utterance(expert, medic, turn-tourniquet(tq1)))
belief(medic, accept-utterance(medic, expert, turn-tourniquet(tq1)))
belief(medic, inform-utterance(medic, expert, stopped(b1)))
belief(medic, acknowledge-utterance(expert, medic, stopped(b1)))
belief(medic, end-dialogue(medic, expert, over-and-out))
```

Table 6. Empirical results from running UMBRA on various dialogues. The rows represent different scenarios or different variations on a scenario. The columns show the number of assumed elements that are true positives (TP), false positives (FP), and false negatives (FN), as well as the precision and recall scores calculated from these values. It also shows the total number of input utterances given (in logical form) in the test instance.

	TP	FP	FN	Inputs	Precision	Recall
Basic						
Scenario 1	380	0	0	25	100.0%	100.0%
Scenario 2	410	8	12	28	98.1%	97.2%
Scenario 3	400	0	0	27	100.0%	100.0%
Scenario 4	390	29	32	28	93.1%	92.4%
Total	*1580*	*37*	*44*	*108*	*97.7%*	*97.3%*
Elided						
No Implicit Speech Acts	310	6	76	19	98.1%	80.3%
Only Medic's Utterances	217	8	174	14	96.3%	55.5%
Only Expert's Utterances	259	38	135	11	87.2%	65.7%
Total	*786*	*52*	*385*	*44*	*93.8%*	*67.1%*

5.1 Reasonable explanations constructed

Our first assertion is that UMBRA constructs appropriate explanations of dialogues. As the upper half of Table 6 shows, the system performed well, generating the target explanations exactly for Scenarios 1 and 3, making 380 and 400 correct assumptions, respectively. It did almost as well with Scenario 2, only making a few spurious inferences as a result of mismatching a clarification subdialogue. In that scenario, the medic asks, "Where should I put it?" and the expert responds, "Below the knee." The medic acknowledges, "Okay." The system interpreted this exchange as a top-level dialogue component instead of a subdialogue punctuating the expert's previous proposal about the tourniquet. It was able to infer a Propose-Accept exchange sans the clarification by assuming that a new instance of the Propose act (with the same content) had occurred immediately before the medic's Accept act.

Scenario 4, the most elaborate dialogue, included both a rejection with a clarification subdialogue and a question the medic could not answer. Nevertheless, UMBRA achieved 93.1% precision and 92.4% recall scores. One of the few bad inferences occurred when the system assumed that the medic had answered the expert's question about the bleeding with an Inform act about his own uncertainty, rather than inferring a more specific rule application meant to encode a response of "I do not know". We believe this reveals inadequate constraints on our dialogue grammar rules – we did not forbid cases in which a regular Inform act encodes an agent's uncertainty about its content – rather than a flaw in the framework *per se*.

We also noted problems when speech acts in exchanges, such as Inform and Acknowledge, are separated by a subdialogue. In such cases, UMBRA sometimes assumes instances of such acts immediately adjacent to their partner, inferring an exchange without the subdialogue component. This occurs in *addition* to making the correct inference, and it takes place because the system has no notion that, in a dialogue, speech acts do not typically overlap and are not repeated at different times. However, these are rules of thumb rather than hard constraints, so revising the system to avoid such failures is not a trivial exercise.

The system also tended to apply domain knowledge widely but gained little for these efforts. Domain content is often relevant to Question acts, whose content must be conceptually linked to its answer, but using it did not reduce UMBRA's errors of omission. Using abduction, it can simply assume a link between two concepts when the knowledge is not available. This ability to assume a conceptual relationship at the domain level led to several instances of faulty reasoning that misaligned questions with their answers.

Despite these occasional problems, the overall results provide evidence that UMBRA can understand reasonably complex dialogues. Although the scenarios were relatively narrow in scope, we consider this to be encouraging preliminary evidence that, given a set of input utterances in logical form, the system generally selects appropriate knowledge and uses it to infer a coherent explanation in terms of the participating agents' beliefs and goals.

5.2 Handling elided utterances

Our second claim was that UMBRA can understand a dialogue from which constituent utterances are elided. A common type of omission in dialogue involves *tacit speech acts* that the conversants implicitly hold to have occurred but that remain unspoken, yet nevertheless play their usual roles in the conversation. Typical examples are the tacit acknowledgment or acceptance of Inform or Propose statements.

We tested UMBRA on an input sequence consisting of the Scenario 1 with its six unspoken Acknowledge and Accept utterances removed. The second half of Table 6 shows the results. The precision score was high, but recall was only 80.3%, partly because the system inferred instances of Propose-Response exchanges without specifying whether they involved acceptances or rejections. As it had evidence for neither, and lacking knowledge that an Accept is the default in such cases, it took the path of least commitment.

Inferring implicit speech acts is reasonably straightforward, but UMBRA can also handle more difficult cases. An extreme example occurs when half of the conversation is missing, as when the listener hears only one side of a telephone conversation. To address the challenging task of restoring such omissions, we took the first scenario, separated the medic's utterances from the expert's, and ran UMBRA on each half separately. The second section of Table 6 shows high precision scores for these explanations: UMBRA makes very few spurious inferences in its attempts to fill in the gaps.[9] The recall scores are substantially lower, at 55.5% for the medic and 65.7% for the expert. Examining the explanations reveals two main reasons for this outcome, other than the least commitment effect already mentioned.

First, UMBRA does not typically work *down* from a rule head in working memory to infer lower-level elements. For instance, when it assumes an Acknowledge act to complete an inference about an Inform-Acknowledge exchange, it does not use that element by default to create an Acknowledge instance. This is a deliberate principle that stems from our system's design as an incremental explanation engine. Often a rule may have multiple decompositions and little is gained by choosing one before acquiring evidence to support that choice. This principle

9. The difference in results for these runs is not surprising, as the dialogue is asymmetric.

holds up less well in dialogue understanding, where there are seldom multiple rule decompositions and there is typically a default choice, such as acceptance over rejection, when they occur.

A second factor is that UMBRA makes some otherwise good inferences with one or more time stamps missing: the lacunae in the dialogue did not always leave enough information to let it determine the exact duration of speech acts or higher-level structures. All such inferences registered as differences from the target explanation. Tolerating these unbound variables would have substantially increased both precision and recall. The effect is more pronounced here than in the previous runs because each elided segment has two time stamps, rather than being instantaneous like the implicit speech acts. Despite these issues, Table 6 reveals that UMBRA still reconstructed two thirds of elided explanations on average, while the number of spurious inferences remained lower than one in ten. On the whole, these results suggest that the system can fill in sizable conversational gaps.

6. Related research

As mentioned earlier, the aim of our research is a computational account of the high level aspects of dialogue understanding. In carrying out this work, we have incorporated and adapted ideas from earlier efforts on language processing and inference.

Our approach to dialogue interpretation has been influenced by four distinct threads of research. The first deals with the notion of speech acts, which goes back to early treatments by Austin (1962) and Searle (1969), and which has received substantial attention in the pragmatics community. In particular, our treatment of the sample dialogue depends on the concept of indirect speech acts, in that we often map utterances onto multiple acts, some of which are implicit. This idea is closely related to work on implicatures (Grice 1975), in which an utterance suggests content beyond its superficial meaning. However, much of the recent effort on implicatures has focused on domain-specific inference, whereas speech acts are primarily domain independent.

Researchers have proposed many taxonomies of speech acts, but we have not adopted any of these in particular for the current work. Rather, we have assumed a few generic categories, such as informing another and asking a question, about which there is little disagreement. We believe it will be possible to revise our framework to incorporate distinctions made by many of these taxonomies. In this approach we have been influenced by Perrault and Allen's (1980) early logical analysis of speech acts, which also dealt with reasoning about other agents' mental states. Our use of domain-independent rules for describing speech acts, which alter the communicating agents' mental states, comes from their work. In a similar vein, Carberry and Lambert (1999) report a dialogue interpretation system that uses 'discourse recipes' that are similar to our speech-act rules. However, we have utilized them to understand joint goal-directed activities, whereas they used them to recognize subdialogues, an ability we have not emphasized in our work.

A second theoretical assumption is that dialogue understanding involves abductive inference, which also has precedents in the literature. For instance, Litman (1985) adopts a similar approach in her work on dialogue processing, although her system incorporated additional linguistic cues that our framework does not. Later work by McRoy and Hirst (1995) also takes an abductive approach to conversational analysis, although they emphasized the process of recovering from misunderstandings.

Ballim and Wilks' (1991) work on belief ascription, including belief ascription for dialogue understanding, is also relevant. We share with their work the idea of using restricted, heuristic reasoning to improve tractability and on-the-fly inference of embedded beliefs. Their approach uses a form of default reasoning in which the system infers that other agents hold the same beliefs as does it unless there is evidence to the contrary. One can view this as a form of abductive reasoning, although they do not describe it this way.

We should also mention Barbella and Forbus' (2011) research on 'analogical dialogue acts', although technically they focused on understanding textual discourse rather than dialogue. For this reason, their analysis focuses on refinements of inform actions related to analogy that often arise in instructional settings. After processing sentences, their system draws inferences using analogical reasoning, which we view as an alternative approach to abductive inference that offers similar functionality. Tomai and Forbus (2009) report another language understanding system that incorporates a form of abduction closer to our own.

The broader literature on abduction includes work by Bullwinkle (1975) and Hobbs et al. (1993), both of whom applied it to sentence-level processing. Our approach has been influenced more directly by Ng and Mooney's (1990) abduction mechanism, which guided search through a space of explanations using a coherence metric. Although they did not focus on dialogue, their system understood relations among sequences of sentences encoded as logical literals. We have also incorporated ideas from Bridewell and Langley's (2011) AbRA, another abduction system that operates in an incremental, data-driven fashion.

Third, our utilization of dialogue grammars builds on early work by Reichman (1981) and by Polanyi and Scha (1984), who used grammatical notations to characterize the structure of conversations. The highlevel rules in Table 2 constitute a simple grammar that encodes information about dialogues' sequential and hierarchical organization. Cohen (1997) has argued that dialogue grammars have inherent limitations, but our approach combines them with meta-level knowledge about the beliefs and goals that are associated with different types of speech acts.

More recent work on discourse understanding uses more sophisticated representations, such as *discourse representation theory* (Kamp & Reyle 1993) and its extensions (e.g. Asher & Lascarides 2003). Research on *discourse obligations* (Traum & Allen 1994; Allen et al. 2001) also encodes knowledge about the sequential structure of dialogues. Our work has abstracted away from language problems like anaphora resolution and entity identification, and we have treated obligations implicitly, which has let us use reasonably simple representations of utterances and dialogue grammars. However, we will need to address these issues when we attempt to automate translation of utterances into logical form and when we examine dialogues in which obligations come into question.

Finally, our overall approach to dialogue understanding borrows heavily from Clark (1996). We have adopted from his framework the view of dialogue as a sequence of actions in which the participants incrementally develop *common ground*, i.e. a set shared beliefs and goals. In addition to augmenting the common ground, as in other discourse models, Clark's dialogue participants engage in the joint task of making sure they mutually understand each other. In his framework, contributing to a dialogue is not an individual participant's action but rather a joint task between the speaker and the listener, where the goal is to achieve mutual understanding of the contribution. Our speech act rules incorporate a similar idea, in that their effects include not only the creation of a belief in the mind of the listener, but also the creation of beliefs in both speaker and listener about the beliefs and the goals that motivated the speech act, as well as beliefs about the occurrence of the speech act itself.

We owe a great intellectual debt to these earlier efforts, many features from which we have combined into a novel, integrated architecture for high-level dialogue understanding. Our work also goes beyond many earlier systems in its construction of explanations that include inferences about the beliefs and goals of participating agents. In other words, it incorporates representations and mechanisms for mind reading as a central part of dialogue interpretation. We have not attempted to model other aspects of this complex process, such as speech recognition or linguistic analysis, but these have already been studied widely.

7. Directions for future work

Although our research to date has produced a promising initial account of dialogue understanding, we must still extend it along a number of fronts. The most obvious next step will involve demonstrating the system's generality by testing it on additional dialogues. Ideally, these should include more complex forms of interaction that require the introduction of knowledge about other types of speech acts and richer dialogue structures. We hypothesize that we can handle these more challenging tasks through the addition of dialogue-level and domain-level rules, and that the basic approach will remain unchanged.

However, we must also extend the framework itself in some important directions. For instance, note that our approach requires as input at least some beliefs about the speech acts in a conversation, and that, following speech act theory, these are quite distinct from the utterances. Although we do not plan to address sentence processing itself, we hope to augment our mechanisms to infer speech acts from the 'literal' meanings of sentences produced by a sentence processor. We believe such inference can also be handled though incremental abduction guided by a dialogue grammar, speech-act rules, and domain-level knowledge.

Additional limitations concern the incremental character of dialogue. One issue involves the granularity of speech acts, which may convey only one aspect of an event or relation. This suggests moving away from complex predicates toward a notation that describes an event (Davidson 1967) as a set of triples that can be communicated separately. Our initial steps in that direction have involved encoding domain-level literals as relational triples so the architecture can access predicates as easily as their arguments (Gabaldon, Langley, & Meadows 2013), but we must still test this approach more thoroughly. A more serious issue is that dialogues are not always as well behaved as our example. There may be unexpected input outside the context of our high-level knowledge, such an interruption act breaking a pattern of speech acts. We should develop mechanisms for relaxed use of dialogue adjacency pairs or otherwise handling ill-formed dialogues gracefully. One response would be to adapt our current abduction process to handle conversations with anomalous utterances.

The dialogue interpretation process may also draw faulty conclusions from which it should recover in light of new evidence. To handle such situations, we must extend UMBRA's monotonic inference mechanism to incorporate methods for belief revision. This process should detect inconsistencies that violate constraints, identify the assumptions that underlie them, and generate an alternative explanation that does not incorporate the questionable elements. The system would adopt this consistent explanation as the new dialogue interpretation and continue, possibly revising it again in the future.

8. Concluding remarks

In the preceding pages, we described the problem of understanding dialogues about joint activities, which we formulated as the inference of participants' mental models of other agents from information about a sequence of speech acts. We introduced a framework for representing mental states, speech acts, and domain-independent rules that relate them. We also described UMBRA, an implemented system that incorporates such knowledge, combined with an incremental form of abductive reasoning, to infer plausible beliefs and goals from observed speech acts.

Testing this system demonstrated that it generates reasonable mental models when provided, incrementally, with all of the speech acts in our target dialogue. We also showed that UMBRA can infer implicit speech acts that are not provided, and that it can even guess one side of a conversation when given only the speech acts on the other side. We argued that the presence of a high-level dialogue grammar is central to this ability, but that domain-level knowledge also plays an important role.

Our approach incorporates ideas from earlier work on pragmatics, dialogue, and abduction, but it combines them in innovative ways to support new capabilities. Although our current implementation is limited in scope, it supports our main theoretical tenets: that dialogue interpretation involves reasoning about the mental states of the participants, that it depends on an incremental form of abductive inference, and that this process operates both over dialogue-level knowledge that is domain independent and over domain-level expertise. Taken together, these claims offer a promising computational account for the interpretation of task-oriented dialogues.

Acknowledgments

This research was supported in part by Grant N00014-09-1-1029 from the Office of Naval Research and in part by a gift from Lockheed Martin Advanced Technology Laboratories. We thank Chitta Baral, Paul Bello, Will Bridewell, Herb Clark, Wende Frost, Tolga Könik, David Stracuzzi, and Richard Weyrauch for discussions that influenced the approach to dialogue understanding we have reported here.

References

Allen, J.F., Byron, D.K., Dzikovska, M., Ferguson, G., Galescu, L., & Stent, A. (2001). Toward conversational human-computer interaction. *AI Magazine*, *22*(4), 27–37.

Asher, N., & Lascarides, A. (2003). *Logics of conversation*. Cambridge, UK: Cambridge University Press.

Austin, J.L. (1962). *How to do things with words*. Cambridge, MA: Harvard University Press. DOI: 10.2307/2269960

Ballim, A., Wilks, Y., & Barnden, J. (1991). Belief ascription, metaphor, and intentional identification. *Cognitive Science*, *15*, 133–171. DOI: 10.1207/s15516709cog1501_4

Barbella, D.M., & Forbus, K.D. (2011). Analogical dialogue acts: Supporting learning by reading analogies in instructional texts. In *Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence* (pp. 1429–1435). San Francisco, CA: AAAI Press.

Bridewell, W., & Isaac, A. (2011). Recognizing deception: A model of dynamic belief attribution. In *Proceedings of the AAAI 2011 Fall Symposium on Advances in Cognitive Systems*. Arlington, VA: AAAI Press.

Bridewell, W., & Langley, P. (2011). A computational account of everyday abductive inference. In *Proceedings of the Thirty-Third Annual Meeting of the Cognitive Science Society*. Boston.

Bullwinkle, C.L. (1975). Picnics, kittens and wigs: Using scenarios for the sentence completion task. In *Proceedings of the Fourth International Joint Conference on Artificial Intelligence* (pp. 383–386).

Cahn, J.E., & Brennan, S.E. (1999). A psychological model of grounding and repair in dialog. In *Proceedings of the AAAI Fall Symposium on Psychological Models of Communication in Collaborative Systems* (pp. 25–33). North Falmouth, MA: AAAI Press.

Carberry, S., & Lambert, L. (1999a). A process model for recognizing communicative acts and modeling negotiation subdialogues. *Computational Linguistics*, *25*, 1–53.
DOI: 10.3115/981967.981992

Clark, H.H. (1996). *Using language*. Cambridge, UK: Cambridge University Press.
DOI: 10.1017/s0022226798217361

Cohen, P. (1997). Dialogue modeling. In R.A. Cole (Ed.), *Survey of the state of the art in human language technology*, Cambridge Studies In Natural Language Processing (pp. 204–210). Cambridge, UK: Cambridge University Press.

Gabaldon, A., Langley, P., & Meadows, B. (2013). Integrating meta-level and domain-level knowledge for interpretation and generation of task-oriented dialogue. In *Proceedings of the Second Annual Conference on Advances in Cognitive Systems* (pp. 171–186). Baltimore, MD: Cognitive Systems Foundation.

Davidson, D. (1967). The logical form of action sentences. In N. Rescher (Ed.), *The logic of decision and action*. Pittsburgh: University of Pittsburgh Press.

Grice, H.P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics 3: Speech acts* (pp. 41–58). New York, NY: Academic Press. DOI: 10.1017/s0022226700005296

Hobbs, J.R., Stickel, M.E., Appelt, D.E., & Martin, P.A. (1993). Interpretation as abduction. *Artificial Intelligence*, *63*, 69–142. DOI: 10.1016/0004–3702(93)90015–4

Kamp, H., & Reyle, U. (1993). *From discourse to logic: Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*. Dordrecht, The Netherlands: Springer.
DOI: 10.1007/978-94-017-1616-1

Litman, D. (1985). Plan recognition and discourse analysis: An integrated approach for understanding dialogues. Ph.D. Thesis, Technical Report 170, Department of Computer Science, University of Rochester.

McRoy, S., & Hirst, G. (1995). The repair of speech act misunderstandings by abductive inference. *Computational linguistics*, *21*, 435–478.

Meadows, B., Langley, P., & Emery, M. (2013a). Seeing beyond shadows: Incremental abductive reasoning for plan understanding. In *Proceedings of the AAAI Workshop on Plan, Activity, and Intent Recognition* (pp. 24–31). Bellevue, WA: AAAI Press.

Meadows, B., Langley, P., & Emery, M. (2013b). Understanding social interactions using incremental abductive inference. In *Proceedings of the Second Annual Conference on Advances in Cognitive Systems* (pp. 39–56). Baltimore, MD: Cognitive Systems Foundation.

Neches, R., Langley, P., & Klahr, D. (1987). Learning, development, and production systems. In D. Klahr, P. Langley, & R. Neches (Eds.), *Production system models of learning and development*. Cambridge, MA: MIT Press.
Ng, H.T., & Mooney, R.J. (1990). On the role of coherence in abductive explanation. In *Proceedings of the Eighth National Conference on Artificial Intelligence* (pp. 337–342). Boston, MA: AAAI Press.
Perrault, C.F., & Allen, J.F. (1980). A plan-based analysis of indirect speech acts. *American Journal of Computational Linguistics*, *6*, 167–182.
Polanyi, R., & Scha, R. (1984). A syntactic approach to discourse semantics. In *Proceedings of the Tenth International Conference on Computational Linguistics* (pp. 413–419). Stanford, CA: Association for Computational Linguistics. DOI: 10.3115/980431.980576
Reichman, R. (1981). *Plain-speaking: A theory and grammar of spontaneous discourse*. Doctoral dissertation, Department of Computer Science, Harvard University, Cambridge, MA.
Sacks, H., Schegloff, E., & Jefferson, G. (1978). A simplest systematics for the organization of turn-taking in conversation. In J. Schenkein (Ed.), *Studies in the organization of conversational interaction*. New York, NY: Academic Press. DOI: 10.1016/b978-0-12-623550-0.50008–2
Searle, J. (1969). *Speech acts: An essay in the philosophy of language*. New York, NY: Cambridge University Press. DOI: 10.1017/s0012217300029073
Thagard, P. R. (1978). The best explanation: Criteria for theory choice. *The Journal of Philosophy*, 75(2), 76–92.
Tomai, E., & Forbus, K.D. (2009) EA NLU: Practical language understanding for cognitive modeling. In *Proceedings of the Twenty-Second International Florida Artificial Intelligence Research Society Conference* (pp. 117–122). Sanibel Island, FL: AAAI Press.
Traum, D.R., & Allen, J.F. (1994) Discourse obligations in dialogue processing. In *Proceedings of the 32nd Annual Meeting on Computational Linguistics* (pp. 1–8). Stroudsburg, PA: Association for Computational Linguistics. DOI: 10.3115/981732.981733
Wielemaker, J., Schrijvers, T., Triska, M., & Lager, T. (2012). SWI-Prolog. *Theory and Practice of Logic Programming*, *12*, 67–96. DOI: 10.1017/s1471068411000494

Authors' addresses

Pat Langley
Silicon Valley Campus
Carnegie Mellon University
Moffett Field
CA 94305 USA

patrick.w.langley@gmail.com

Ben Meadows
Computer Science Department
University of Auckland
Private Bag 92019
Auckland, New Zealand

bmea011@aucklanduni.ac.nz

Alfredo Gabaldon
Silicon Valley Campus
Carnegie MellonUniversity
Moffett Field
CA 94305 USA

alfredo.gabaldon@sv.cmu.edu

Richard Heald
Computer Science Department
University of Auckland
Private Bag 92019
Auckland, New Zealand

rhea335@aucklanduni.ac.nz

Authors' biographies

Pat Langley serves as Distinguished Scientist at Carnegie Mellon University's Silicon Valley Campus and as Professor of Computer Science at the University of Auckland. His research addresses issues in cognitive architectures, dialogue systems, social cognition, and computational scientific discovery.

Ben Meadows is a research assistant at the University of Auckland. His work in cognitive systems has focused on social reasoning and models of explanation and understanding.

Alfredo Gabaldon is Research Scientist at Carnegie Mellon University, Silicon Valley Campus. His research interests are in artificial intelligence, with a focus on knowledge representation and automated reasoning.

Richard Heald is a research assistant at the University of Auckland. His work in cognitive systems has focused on models of question answering and social explanation.