Speech Acts and Rationality

Philip R. Cohen
Artificial Intelligence Center
SRI International
and
Center for the Study of Language and Information
Stanford University

Hector J. Levesque
Department of Computer Science
University of Toronto*

1 Abstract

This paper derives the basis of a theory of communication from a formal theory of rational interaction. The major result is a demonstration that illocutionary acts need not be primitive, and need not be recognized. As a test case, we derive Searle's conditions on requesting from principles of rationality coupled with a Gricean theory of imperatives. The theory is shown to distinguish insincere or nonserious imperatives from true requests. Extensions to indirect speech acts, and ramifications for natural language systems are also briefly discussed.

2 Introduction

The unifying theme of much current pragmatics and discourse research is that the coherence of dialogue is to be found in the interaction of the conversants' *plans*. That is, a speaker is regarded as planning his utterances to achieve his goals, which may involve influencing a hearer by the use of communicative or "speech" acts. On receiving an utterance realizing such an action, the hearer attempts to infer the speaker's goal(s) and to understand how the utterance furthers them. The hearer then adopts new goals (e.g., to respond to a request, to clarify the previous speaker's utterance or goal) and plans his own utterances to achieve those. A conversation ensues. [1]

This view of language as purposeful action has pervaded Computational Linguistics research, and has resulted in numerous prototype systems [1, 2, 3, 5, 9, 25, 27]. However, the formal foundations underlying these systems have been unspecified or underspecified. In this state of affairs, one cannot characterize what a system *should* do independently from what it does.

This paper begins to rectify this situation by presenting a formalization of rational interaction, upon which is erected the beginnings of a theory of communication and speech acts. Interaction is derived from principles of rational action for individual agents, as well as principles of belief and goal adoption among agents. The basis of a theory of purposeful communication thus emerges as a consequence of principles of action.

*Fellow of the Canadian Institute for Advanced Research.

[1]This research was made possible in part by a gift from the Systems Development Foundation, and in part by support from the Defense Advanced Research Projects Agency under Contract N00039-84-K-0078 with the Naval Electronic Systems Command. The views and conclusions contained in this document are those of the authors and should not be interpreted as representative of the official policies, either expressed or implied, of the Defense Advanced Research Projects Agency or the United States Government. Much of this research was done when the second author was employed at the Fairchild Camera and Instrument Corp.

2.1 Speech Act Theory

Speech act theory was originally conceived as part of action theory. Many of Austin's [4] insights about the nature of speech acts, felicity conditions, and modes of failure apply equally well to non-communicative actions. Searle [26] repeatedly mentions that many of the conditions he attributes to various illocutionary acts (such as requests and questions) apply more generally to non-communicative action. However, researchers have gradually lost sight of their roots. In recent work [28] illocutionary acts are formalized, and a logic is proposed, in which properties of IA's (e.g., "preparatory conditions" and "modes of achievement") are primitively stipulated, rather than derived from more basic principles of action. We believe this approach misses significant generalities. This paper shows how to derive properties of illocutionary acts from principles of rationality, updating the formalism of [10].

Work in Artificial Intelligence provided the first formal grounding of speech act theory in terms of planning and plan recognition, culminating in Perrault and Allen's [22] theory of indirect speech acts. Much of our research is inspired by their analyses. However, one major ingredient of their theory can be shown to be redundant [10] illocutionary acts. All the inferential power of the recognition of their illocutionary acts was already available in other "operators". Nevertheless, the natural language systems based on this approach [1, 5] always had to recognize which illocutionary act was performed in order to respond to a user's utterance. Since the illocutionary acts were unnecessary for achieving their effects, so too was their recognition.

The stance that illocutionary acts are not primitive, and need not be recognized, is a liberating one. Once taken, it becomes apparent that many of the difficulties in applying speech act theory to discourse, or to computer systems, stem from taking these acts too seriously — i.e., too primitively.

3 Form of the argument

We show that illocutionary acts need not be primitive by deriving Searle's conditions on requesting from an independently-motivated theory of action. The realm of communicative action is entered following Grice [13] — by postulating a correlation between the utterance of a sentence with a certain syntactic feature (e.g., its dominant clause is an imperative) and a complex propositional attitude expressing the speaker's goal. This attitude becomes true as a result of uttering a sentence with that feature. Because of certain general principles governing beliefs and goals, other causal consequences of the speaker's having the expressed goal can be derived. Such derivations will be "summarized" as lemmas of the form "If (conditions) are true, then any action making (antecedent) true also makes (consequent) true." These lemmas will be used to characterize illocutionary acts, though they are not themselves acts. For example, the lemma called **REQUEST** will characterize a derivation that shows how a hearer's knowing that the speaker has certain goals can cause the hearer to act. The conditions licensing that chain will be collected in the **REQUEST** lemma, and will be shown to subsume those stipulated by Searle [26] as felicity conditions. However, they have been derived here from first principles, and without the need for a primitive action of requesting.

The benefits of this approach become clearer as other illocutionary acts are derived. We have derived a characterization of the speech act of informing, and have used it in deriving the speech act of questioning. The latter derivation also allows us to distinguish real questions from teacher/student questions, and rhetorical questions. However, for brevity, the discussion of these speech acts has been omitted.

Indirect speech acts can be handled within the framework, although, again, we cannot present the analyses here. Briefly, axioms similar to those of Perrault and Allen [22] can be supplied enabling one to reason that an agent has a goal that q, given that he also has a goal p. When the p's and q's are themselves goals of the hearer (i.e., the speaker is trying to get the hearer to do something), then we can derive a set of lemmas for indirect requests. Many of these indirect request lemmas correspond to what have been called "short-circuited" implicatures, which, it was suggested [21] underlie the processing of utterances of the form "Can you do X?", "Do you know y?", etc. Lemma formation and lemma application thus provide a familiar model of short-circuiting. Furthermore, this approach shows how one can use general purpose reasoning in concert with conventionalized forms (e.g., how one can reason that "Can you reach the salt" is a request to pass the salt), a problem that has plagued most theories of speech acts.

The plan for the paper is to construct a formalism based on a theory of action that is sufficient for characterizing a request. Most of the work is in the theory of action, as it should be.

4 The Formalism

To achieve these goals we need a carefully worked out (though perhaps incomplete) theory of rational action and interaction. The theory will be expressed in a logic whose model theory is based (loosely) on a possible-worlds semantics. We shall propose a logic with four primary modal operators -- **BEL**ief, **BMB**, **GOAL**, and **AFTER**. With these, we shall characterize what agents need to know to perform actions that are intended to achieve their goals. The agents do so with the knowledge that other agents operate similarly. Thus, agents have beliefs about other's goals, and they have goals to influence others' beliefs and goals. The integration of these operators follows that of Moore [20], who analyzes how an agent's knowledge affects and is affected by his actions, by meshing a possible-worlds model of knowledge with a situation calculus model of action [18]. By adding **GOAL**, we can begin to talk about an agent's plans, which can include his plans to influence the beliefs and goals of others.

Intuitively, a model for these operators includes courses of events (i.e., sequences of primitive acts) [2] that characterize what has happened. Courses of events (*c.o.e.*'s) are paths through a tree of possible future primitive acts, and after any primitive act has occurred, one can recover the course of events that led up to it. C.o.e.'s can also be related to one another via accessiblity relations that partake in the semantics of **BEL** and **GOAL**. Further details of this semantics must await our forthcoming paper [17].

As a general strategy, the formalism will be too strong. First, we have the usual consequential closure problems that plague possible-worlds models for belief. These, however, will be accepted for the time being. Second, the formalism will describe agents as satisfying certain properties that might generally be true, but for which there might be exceptions. Perhaps a process of non-monotonic reasoning could smooth over the exceptions, but we will not attempt to specify such reasoning here. Instead, we assemble a set of basic principles and examine their consequences for speech act use. Third, we are willing to live with the difficulties of the situation calculus model of action — e.g., the lack of a way to capture true parallelism, and the frame problem. Finally, the formalism should be regarded as a description or specification *of* an agent, rather than one that any agent could or should use.

Our approach will be to ground a theory of communication in a theory of rational interaction, itself supported by a theory of rational action, which is finally grounded in mental states. Accordingly, we first need to describe the behavior of **BEL**, **BMB**, **GOAL** and **AFTER**. Then, these operators will be combined to describe how agents' goals and plans influence their actions. Then, we characterize how having beliefs about the beliefs and goals of others can affect one's own beliefs and goals. Finally, we characterize a request.

To be more specific, here are the primitives that will be used, with a minimal explanation.

4.1 Primitives

Assume **p, q**, ... are schema variables ranging over wffs, and **a, b** ... are schematic variables ranging over acts. Then the following are wffs.

4.1.1 Wffs

~p
(p ∨ q)
(**AFTER** a p) - p is true in all courses of events that obtain from act a's happening [3], (if a denotes a halting act).
(**DONE** a) - The event denoted by a has *just* happened.
(**AGT** a x) - Agent x is the only agent of act a
a ≤ b - Act a precedes act b in the current course of events.
∃z p where p contains a free occurrence of variable *z*.
x = y
True, *False*
(**BEL** x p) - p *follows from* x's beliefs.
(**GOAL** x p) - p *follows from* x's goals.
(**BMB** x y p) - p *follows from* x's beliefs about what is mutually believed by x and y.

[2] For this paper, the only events that will be considered are primitive acts.

[3] That is, p is true in all c.o.e.'s resulting from concatenating the current c.o.e. with the c.o.e. denoted by a.

4.1.2 Action Formation

If **a, b, c, d** range over sequences of primitive acts, and **p** is a wff, then the following are complex act descriptions:

a;b — sequential action

a | b — non-deterministic choice (a or b) action

p? — action of positively testing **p**.

(IF p a b) — conditional action $\stackrel{\text{def}}{=}$ **(p?;a) | (~p?;b)**, as in dynamic logic.

(UNTIL p a) — iterative action $\stackrel{\text{def}}{=}$ **(~p;a)*;~p?** (again, as in dynamic logic).

The meta-symbol '$\vdash$' will prefix formulas that are theorems, i.e., that are derivable. Properties of the formal system that will be assumed to hold will be termed *Propositions*. Propositions will be both formulas that should always be valid, for our forthcoming semantics, and rules of inference that should be sound. No attempt to prove or validate these propositions here, but we do so in [17].

4.2 Properties of Acts

We adopt the usual axioms characterizing how complex actions behave under **AFTER**, as treated in a dynamic logic (e.g., [20]) namely,

Proposition 1 *Properties of complex acts* —

(AFTER a;b p) ≡ (AFTER a (AFTER b p)).
(AFTER a|b p) ≡ (AFTER a p) ∧ (AFTER b p).
(AFTER p? q) ≡ p ∧ q.

AFTER and **DONE** will have the following additional properties:

Proposition 2 **∀ act (AFTER act (DONE x act))** [4]

Proposition 3 **∀a [(DONE (AFTER a p)?;a) ⊃ p]**

Proposition 4 *If* $\vdash \alpha \supset \beta$ *then*
$\vdash$ **∀a** **(DONE α?;a) ⊃ (DONE β?;a)**

Proposition 5 **p ≡ (DONE p?)**

Proposition 6 **(DONE [(p ⊃ q) ∧ p]?) ⊃ (DONE q?)**

Our treatment of acts requires that we deal somehow with the "frame problem" [18]. That is, we must characterize not only what changes as a result of doing an action, but also what does not change. To approach this problem, the following notation will be convenient:

Definition 1 **(PRESERVES a p)** $\stackrel{\text{def}}{=}$ **p ⊃ (AFTER a p)**

Of course, all theorems are preserved.

Temporal concepts are introduced with **DONE** (for past happenings) and ◇ (read "eventually"). To say that **p** was true at some point in the past, we use **∃a (DONE p?;a)**. ◇ is to be regarded in the "branching time" sense [11], and will be defined more rigorously in [17]. Essentially, **◇p** is true iff for all infinite extensions of any course of events there is a finite prefix satisfying p. **◇p** and **◇~p** are jointly satisfiable. Since **◇p** starts "now", the following property is also true,

Proposition 7 $\vdash$ **p ⊃ ◇p**

Also, we have the following rule of inference:

Proposition 8 *If* $\vdash \alpha \supset \beta$ *then* $\Diamond(\alpha \vee p) \supset \Diamond(\beta \vee p)$

4.3 The Attitudes

Neither **BEL**, **BMB**, nor **GOAL** characterize what an agent actively believes, mutually believes (with someone else), or has as a goal, but rather what is *implicit* in his beliefs, mutual beliefs, and goals. [5] That is, these operators characterize what the world would be like if the agent's beliefs and mutual beliefs were true, and if his goals were made true. Importantly, we do not include an operator for wanting, since desires need not be consistent. We assume that once an agent has sorted out his possibly inconsistent desires in deciding what he wishes to achieve, the worlds he will be striving for are consistent. Conversely, recognition of an agent's plans need not consider that agent's possibly inconsistent desires. Furthermore, there is also no explicit operator for intending. If an agent intends to bring about p, the agent is usually regarded as also being able to bring about p. By using **GOAL**, we will be able to reason about the end state the agent is aiming at separately from our reasoning about his ability to achieve that state.

For simplicity, we assume the usual Hintikka axiom schemata for **BEL** [15], and we introduce **KNOW** by definition:

Definition 2 **(KNOW x p)** $\stackrel{\text{def}}{=}$ **p ∧ (BEL x p)**

4.3.1 Mutual Belief

Human communication depends crucially on what is mutually believed [1, 6, 7, 9, 22, 23, 24]. We do not use the standard definitions, but employ **(BMB y x p)**, which stands for y's belief that it is mutually believed between y and x that p. **(BMB y x p)** is true iff **(BEL y [p ∧ (BMB x y p)])**. [6] **BMB** has the following properties:

Proposition 9 **(BMB y x p∧q) ≡ (BMB y x p) ∧ (BMB y x q)**

Proposition 10 **(BMB y x p⊃q) ⊃ ((BMB y x p) ⊃ (BMB y x q))**

Proposition 11 *If* $\vdash \alpha \supset \beta$ *then* $\vdash$**(BMB y x α) ⊃ (BMB y x β)**

Also, we characterize mutual knowledge as:

Definition 3 **(MK x y p)** $\stackrel{\text{def}}{=}$ **p ∧ (BMB x y p) ∧ (BMB y x p)**[7]

[4] **(AFTER t (DONE t))**, where **t** is term denoting a primitive act (or a sequence of primitive acts), is not always true since an act may change the values of terms (e.g., an election changes the value of the term **(PRESIDENT U.S.)**)

[5] For an exploration of the issues involved in explicit vs. implicit belief, see [16].

[6] Notice that **(BMB y x p) ≢ (BMB x y p)**.

[7] This definition is not entirely correct, but is adequate for present purposes.

4.3.2 Goals

For GOAL, we have the following properties:

Proposition 12 (GOAL x (GOAL x p)) $\supset$ (GOAL x p)

If an agent thinks he has a goal, then he does.

Proposition 13 (BEL x (GOAL x p)) $\equiv$ (GOAL x p)

Proposition 14 (GOAL x p) $\wedge$ (GOAL x p$\supset$q) $\supset$ (GOAL x q)[6]

The following two derived rules are also useful:

Proposition 15 *If* $\vdash \alpha \supset \beta$ *then* $\vdash$(GOAL x α) $\supset$ (GOAL x β)

Proposition 16 *If* $\vdash \alpha \wedge \beta \supset \gamma$ *then* $\vdash$(BMB y x (GOAL x α)) $\wedge$ (BMB y x (GOAL x β)) $\supset$ (BMB y x (GOAL x γ))

More properties of GOAL follow.

4.4 Attitudes and Rational Action

Next, we must characterize how beliefs, goals, and actions are related. The interaction of **BEL** and **AFTER** will be patterned after Moore's analysis [20]. In particular, we have:

Proposition 17 $\forall$ x, act (AGT a x) $\supset$ (AFTER act (KNOW x (DONE act)))

Agents know what they have done. Moreover, they think certain effects of their own actions are achieved:

Proposition 18 (BEL x (RESULT x a p)) $\supset$ (RESULT x a (BEL x p)), *where*

Definition 4 (RESULT x a p) $\stackrel{def}{=}$ (AFTER a p) $\wedge$ (AGT a x)

The major addition we have made is GOAL, which interacts tightly with the other operators.

We will say a rational agent only adopts goals that are achievable, and accepts as "desirable" those states of the world that are inevitable. To characterize inevitabilities, we have

Definition 5 (ALWAYS p) $\stackrel{def}{=}$ $\forall$a (AFTER a p)

This says that *no matter what* happens, p is true. Clearly, we want

Proposition 19 *If* $\vdash\alpha$ *then* $\vdash$ (BEL x (ALWAYS α))

That is, theorems are believed to be always true.

Another property we want is that no sequence of primitive acts is forever ruled out from happening.

Proposition 20 $\vdash$ $\forall$a (ACT a) $\supset$ ~(ALWAYS ~(DONE a)), where (ACT a) $\stackrel{def}{=}$ ~(AFTER a ~(DONE a))

One important variant of ALWAYS is (ALWAYS x p) (relative to an agent), which indicates that no matter what *that agent* does, p is true. The definition of this version is:

Definition 6 (ALWAYS x p) $\stackrel{def}{=}$ $\forall$a (RESULT x a p)

A useful instance of ALWAYS is (ALWAYS p$\supset$q) in which no matter what happens, p still implies q. We can now distinguish between p $\supset$ q's being logically valid, its being true in all courses of events, and its merely being true after some event happens.

[6] Notice that if p$\supset$q is true (or even believed) but (GOAL x p$\supset$q) is not true, we should not reach this conclusion since some act could make it false.

4.4.1 Goals and Inevitabilities

What an agent believes to be inevitable is a goal (he accepts what he cannot change).

Proposition 21 (BEL x (ALWAYS p)) $\supset$ (GOAL x p)

and conversely (almost), agents do not adopt goals that they believe to be impossible to achieve —

Proposition 22 *No futility* — (GOAL x p) $\supset$ ~(BEL x (ALWAYS ~p))

This gives the following useful lemma:

Lemma 1 *Inevitable Consequences*
(GOAL x p) $\wedge$ (BEL x (ALWAYS p$\supset$q)) $\supset$ (GOAL x q)

Proof: By Proposition 21, if an agent believes p$\supset$q is always true, he has it as a goal. Hence by Proposition 14, q follows from his goals.

This lemma states that if one's goal is a c.o.e. in which p holds, and if one thinks that no matter what happens, p$\supset$q, then one's goal is a c.o.e. in which q holds. Two aspects of this property are crucially important to its plausibility. First, one must keep in mind the "follows from" interpretation of our propositional attitudes. Second, the key aspect of the connection between p and q is that *no one* can achieve p without achieving q. If someone could do so, then q need not be true in a c.o.e. that satisfies the agent's goals.

Now, we have the following as a lemma that will be used in the speech act derivations:

Lemma 2 *Shared Recognition*

(BMB y x (GOAL x p)) $\wedge$
(BMB y x (BEL x (ALWAYS p$\supset$q))) $\supset$
(BMB y x (GOAL x q))

The proof is a straightforward application of Lemma 1 and Propositions 9 and 10.

4.4.2 Persistent goals

In this formalism, we are attempting to capture a number of properties of what might be called "intention" without postulating a primitive concept for "intend". Instead, we will combine acts, beliefs, goals, and a notion of commitment built out of more primitive notions.

To capture one grade of commitment than an agent might have towards his goals, we define a persistent goal, P-GOAL, to be one that the agent will not give up until he thinks it has been satisfied, or until he thinks he cannot achieve it.

Now, in order to state constraints on c.o.e.'s we define:

Definition 7 (PREREQ x p q) $\stackrel{def}{=}$
$\forall c$ (RESULT x c q) $\supset$ $\exists$ a (a $\preceq$ c) $\wedge$ (RESULT x a p)

This definition states that p is a prerequisite for x's achieving q if *all* ways for x to bring about q result in a course of events in which p has been true. Now, we are ready for persistent goals:

Definition 8 (P-GOAL x p) $\stackrel{def}{=}$
(GOAL x p) $\wedge$
[PREREQ x ((BEL x p) $\vee$
(BEL x (ALWAYS x ~p)))
~(GOAL x p)]

Persistent goals are ones the agent will replan to achieve if his earlier attempts to achieve it fail to do so. Our definition does *not* say that an agent *must* give up his goal when he thinks it is satisfied, since goals of maintenance are allowed. All this says is that somewhere along the way to giving up the persistent goal, the agent had to think it was true (or believe it was impossible for him to achieve).

Though an agent may be persistent, he may be foolishly so because he has no competence to achieve his goals. We characterize competence below.

4.4.3 Competence

People are sometimes experts in certain fields, as well as in their own bodily movements. For example, a competent electrician will form correct plans to achieve world states in which "electrical" states of affairs obtain. Most adults are competent in achieving world states in which their teeth are brushed, etc. We will say an agent is **COMPETENT** with respect to p if, whenever he thinks p will true after some action happens, he is correct:

Definition 9 (COMPETENT x p) $\stackrel{def}{=}$
$\forall$a (BEL x (AFTER x p)) $\supset$ (AFTER a p)

One property of competence we will want is:

Proposition 23 $\forall$x, a (AGT x a) $\supset$
(ALWAYS (COMPETENT x (DONE x a))), *where*

Definition 10 (DONE x a) $\stackrel{def}{=}$ (DONE a) $\wedge$ (AGT a x)

That is, any person is always competent to do the acts of which he is the agent. [9] Of course, he is not always competent to achieve any particular effect.

Finally, given all these properties we are ready to describe rational agents.

4.5 Rational Agents

Below are properties of ideally rational agents who adopt persistent goals.

First, agents are careful; they do not knowingly and deliberately make their persistent goals impossible for them achieve.

Proposition 24 (DONE x act) $\supset$ (DONE x p?;act), where
p $\stackrel{def}{=}$ (P-GOAL x q) $\supset$ ~(BEL x (AFTER act
(ALWAYS x ~p))) $\vee$
~(GOAL x (DONE x act)) [10]

In other words, no deliberately shooting onesself in the foot.

Now, agents are cautious in adopting persistent goals, since they must eventually come to some decision about their feasibility. We require an agent to either come up with a "plan" to achieve them — a belief of some act (or act sequence) that it achieves the persistent goal — or to believe he cannot bring the goal about. That is, agents do not adopt persistent goals they could never give up. The next Proposition will characterize this property of **P-GOAL**.

But, even with a correct plan and a persistent goal, there is still the possibility that the competent agent never executes the plan in the right circumstances — some other agent has changed the circumstances, thereby making the plan incorrect. If the agent is competent, then if he formulates another plan, it will be correct for the new circumstances. But again, the world could change out from under him. Now, just as with operating systems, we want to say that the world is "fair" - the agent will eventually get a chance to execute his plans. This property is also characterized in the following Proposition:

Proposition 25 *Fair Execution — The agent will eventually form a plan and execute it, believing it achieves his persistent goal in circumstances he believes to be appropriate for its success.*

$\forall$ x (P-GOAL x q) $\supset$
$\Diamond$[$\exists$ act' (DONE x p?;act')] $\vee$
[BEL x (ALWAYS x ~q)]),
where p $\stackrel{def}{=}$ (BEL x (RESULT x act' q))

We now give a crucial theorem:

Theorem 1 *Consequences of a persistent goal — If someone has a persistent goal of bringing about* p, *and bringing about* p *is within his area of competence, then eventually either* p *becomes true or he will believe there is nothing that can be done to achieve* p

$\forall$ y (P-GOAL y p) $\wedge$ (ALWAYS (COMPETENT y p)) $\supset$
$\Diamond$ (p $\vee$ (BEL y (ALWAYS y ~p)))

Proof sketch:
Since the agent has a persistent goal, he eventually will either find and execute a plan, or will believe there is nothing he can do to achieve the goal. Since he is competent with respect to p, the plans he forms will be correct. Since his plan **act'** is correct, and since any other plans he forms for bringing about p are also correct, and since the world is "fair", eventually either the agent executes his correct plan, making p true, or the agent comes to believe he cannot achieve p. A more rigorous proof can be found in the Appendix.

This theorem is a major cornerstone of the formalism, telling us when we can conclude $\Diamond$p, given a plan and a goal, and is used throughout the speech act analyses. If an agent who is not **COMPETENT** with respect to p adopts p as a persistent goal, we cannot conclude that eventually either p will be true (or the agent will think he cannot bring it about), since the agent could forever create incorrect plans. If the goal is not persistent, we also cannot conclude $\Diamond$p since the agent could give it up without achieving it.

The use of $\Diamond$ opens the formalism to McDermott's "Little Nell" paradox [19]. [11] In our context, the problem arises as follows: First, since an agent has a persistent goal to achieve p,

[9] Because of Proposition 2, all Proposition 23 says is that if a competent agent believes his own primitive act halts, it will.

[10] Notice that it is crucial that p be true in the same world in which the agent does act, hence the use of "p?;act".

[11] Little Nell is tied to the railroad tracks, and will be mashed by the next train. Dudley Doright is planning to save her. McDermott claims that, according to various AI theories of planning, he never will, even though he always knows just what to do.

and we assume here he is always competent with respect to p, $\Diamond$p is true. But, when p is of the form $\Diamond$q (e.g., $\Diamond$(SAVED LITTLE-NELL)), $\Diamond\Diamond$q is true, so $\Diamond$q is true as well. Let us assume the agent knows all this. Hence, by the definition of **P-GOAL**, one might expect the agent to give up his persistent goal that $\Diamond$q, since it is already satisfied!

On the other hand, it would appear that Proposition 25 is sufficient to prevent the agent from giving up his goal too soon, since it states that the agent with a persistent goal must act on it, and, moreover, the definition of **P-GOAL** does not require the agent to give up his goal immediately. For persistent goals to achieve $\Diamond$q, within someone's scope of competence, one might think the agent need "only" *maintain* $\Diamond$q as a goal, and then the other properties of rationality force the agent to perform a primitive act.

Unfortunately, the properties given so far do not yet rule out Little Nell's being mashed, and for two reasons. First, **NIL** denotes a primitive act — the empty sequence. Hence, doing it would satisfy Proposition 25, but the agent never does anything substantive. Second, doing anything that does not affect q also satisfies Proposition 25, since after doing the unrelated act, $\Diamond$q is still true. We need to say that the agent eventually acts on q! To do so, we have the following property:

Proposition 26 (**P-GOAL** y $\Diamond$q) $\supset$
$\Diamond$[(**P-GOAL** y q) $\vee$
(**BEL** y (**ALWAYS** y $\sim$q))].

That is, eventually the agent will have the persistent goal that q, and by Proposition 25, will act on it. If he eventually comes to believe he cannot bring about q, he eventually comes to believe he cannot bring about eventually q as well, allowing him to give up his persistent goal that eventually q.

4.6 Rational Interaction

This ends our discussion of single agents. We now need to characterize rational interaction sufficiently to handle a simple request. First, we discuss cooperative agents, and then the effects of uttering sentences.

4.6.1 Properties of Cooperative Agents

We describe agents as sincere, helpful, and more knowledgeable than others about the truth of some state of affairs. Essentially, these concepts capture (quite simplistic) constraints on influencing someone else's beliefs and goals, and on adopting the beliefs and goals of someone else as one's own. More refined versions are certainly desirable. Ultimately, we expect such properties of cooperative agents, as embedded in a theory of rational interaction, to provide a formal description of the kinds of conversational behavior Grice [14] describes with his "conversational maxims".

First, we will say an agent is **SINCERE** with respect to p if whenever his goal is to get someone else to *believe* p, his goal is in fact to get that person to *know* p.

Definition 11 (**SINCERE** x p) $\stackrel{def}{=}$
(**GOAL** x (**BEL** y p)) $\supset$ (**GOAL** x (**KNOW** y p))

An agent is **HELPFUL** to another if he adopts as his own persistent goal another agent's goal that he eventually do something (provided that potential goal does not conflict with his own).

Definition 12 (**HELPFUL** x y) $\stackrel{def}{=}$
$\forall$a (**BEL** x (**GOAL** y $\Diamond$(**DONE** y a))) $\wedge$
$\sim$(**GOAL** x $\sim$(**DONE** x a)) $\supset$
(**P-GOAL** x (**DONE** x a))

Agent x thinks agent y is more **EXPERT** about the true of p than x if he always adopts x's beliefs about p as his own.

Definition 13 (**EXPERT** y x p) $\stackrel{def}{=}$
(**BEL** x (**BEL** y p)) $\supset$ (**BEL** x p)

4.6.2 Uttering Sentences with Certain "Features"

Finally, we need to describe the effects of uttering sentences with certain "features" [14], such as mood. In particular, we need to characterize the results of uttering imperative, interrogative, and declarative sentences [12] Our descriptions of these effects will be similar to Grices's [13] and to Perrault and Allen's [22] "surface speech acts". Many times, these sentence forms are not used literally to perform the corresponding speech acts (requests, questions, and assertions).

The following is used to characterize uttering an imperative:

Proposition 27 *Imperatives:*

$\forall$ x y (**MK** x y (**ATTEND** y x)) $\supset$
(**RESULT** x [**IMPER** x y "do y act"]
(**BMB** y x
(**GOAL** x
(**BEL** y
(**GOAL** x
(**P-GOAL** y (**DONE** y act)
)))))))

The act [**IMPER speaker hearer** 'p] stands for "make p true". Proposition 27 states that if it is mutually known that y is attending to x, [13] then the result of uttering an imperative to y to make it the case that y has done action act is that y thinks it is mutually believed that the speaker's goal is that y should think his goal is for y to form the persistent goal of doing act.

We also need to assert that **IMPER** preserves sincerity about the speaker's goals and helpfulness. These restrictions could be loosened, but maintaining them is simpler.

Proposition 28 (**PRESERVES** [**IMPER** x y "do y act"]
(**BMB** y x (**SINCERE** y (**GOAL** y p))))

Proposition 29 (**PRESERVES** [**IMPER** x y "do y act"]
(**HELPFUL** y x))

All Gricean "feature"-based theories of communication need to account for cases in which a speaker uses an utterance with a feature, but does not have the attitudes (e.g., beliefs, and goals)

[12] However, we can only present the analysis of imperatives here.

[13] If it is not mutually known that y is attending, for example, if the speaker is not speaking to an audience, then we do not say what the result of uttering an imperative is.

usually attributed to someone uttering sentences with that feature. Thus, the attribution of the attitudes needs to be context-dependent. Specifically, proposition 28 needs to be weak enough to prevent nonserious utterances such as "go jump in the lake" from being automatically interpreted as requests even though the utterance is an imperative. On the other hand, the formula must be strong enough that requests are derivable.

5 Deriving a Simple Request

In making a request, the speaker is trying to get the hearer to do an act. We will show how the speaker's uttering an imperative to do the act leads to its eventually being done. What we need to prove is this:

Theorem 2 *Result of an Imperative* —
(DONE [(MK x y (ATTEND y x)) $\wedge$
(BMB y x
(SINCERE x
(GOAL x
(P-GOAL y (DONE y act)))))$\wedge$
(HELPFUL y x)]?;
[IMPER x y "do y act"]) $\supset$
$\Diamond$**(DONE y act)**

We will give the major steps of the proof in Figure 1, and point to their justifications. The full-fledged proofs are left to the energetic reader. All formulas preceded by a • are supposed to be true just prior to performing the **IMPER**, are preserved by it, and thus are implicitly conjoined to formulas 2 - 9. By their placement in the proof, we indicate where they are necessary for making the deductions.

Essentially, the proof proceeds as follows:

If it is mutually known that y is attending to x, and y thinks it is mutually believed that the •-conditions hold, then x's uttering an imperative to y to do some action results in formula (2). Since it is mutually believed x is sincere about his goals, then (3) it is mutually believed his goal truly is that y form a persistent goal to do the act. Since everyone is always competent to do acts of which they are the agent, (4) it is mutually believed that the act will eventually be done, or y will think it is forever impossible to do. But since no halting act is forever impossible to do, it is (5) mutually believed that x's goal is that y eventually do it. Hence, (6) y thinks x's goal is that y eventually do the act. Now, since y is helpfully disposed towards x, and has no objections to doing the act, (7) y takes it on as a persistent goal. Since he is always competent about doing his own acts, (8) eventually it will be done or he will think it impossible to do. Again, since it is not forever impossible, (9) he will eventually do it.

We have shown how the performing of an imperative to do an act leads to the act's eventually being done. We wish to create a number of lemmas from this proof (and others like it) to characterize illocutionary acts.

6 Plans and Summaries

6.1 Plans

A *plan* for agent "x" to achieve some goal "q" is an action term "a" and two sequences of wffs "p_0", "p_1", ... "p_k" and "q_0", "q_1", ... "q_k" where "q_k" is "q" and satisfying

1. $\vdash$ **(BEL x** $(p_0 \wedge p_1 \wedge \ldots \wedge p_k) \supset$
   (RESULT x a $q_0 \wedge p_1 \wedge \ldots \wedge p_k$ **)))**

2. $\vdash$ **(BEL x (ALWAYS** $(p_i \wedge q_{i-1}) \supset q_i$**)))** $i=1,2,\ldots k$

In other words, given a state where "x" believes the "p_i", he will believe that if he does "a" then "q_0" will hold and moreover, given that the act preserves p_i, and he believes his making "q_{i-1}" true in the presence of p_i will also make "q_i" true. Consequently, a plan is a special kind of proof that

$\vdash$ **(BEL x** $((p_0 \wedge \ldots \wedge p_k) \supset$ **(RESULT x a q)))**

and therefore, since

(BEL x p) $\supset$ **(BEL x (BEL x p))**

and

(BEL x (p $\supset$ **q))** $\supset$ **((BEL x p)** $\supset$ **(BEL x q))**, are axioms of belief, a plan is a proof that

$\vdash$ **(BEL x** $(p_0 \wedge \ldots \wedge p_k)$**)** $\supset$ **(BEL x (RESULT x a q))**

Among the corollaries to a plan are

$\vdash$ **(BEL x (** $(p_0 \wedge \ldots \wedge p_i) \supset$ **(RESULT x a** q_i**)))** $i=1,\ldots k$

and

$\vdash$ **(BEL x (** $(p_i \wedge \ldots \wedge p_j) \supset$ **(ALWAYS** $q_{i-1} \supset q_j$**)))** $i=1,\ldots k$ $j=i,\ldots k$

There are two main points to be made about these corollaries. First of all, since they are theorems, the implications can be taken to be believed by the agent "x" in every state. In this sense, these wffs express general methods believed to achieve certain effects provided the assumptions are satisfied. The second point is that these corollaries are in precisely the form that is required in a plan and therefore can be used as justification for a step in a future plan in much the same way a lemma becomes a single step in the proof of a theorem.

6.2 Summaries

We therefore propose a notation for describing many steps of a plan as a single summarizing operator. A *summary* consists of a name, a list of free variables, a distinguished free variable called the *agent* of the summary (who will always be listed first), an *Effect* which is a wff, a optional *Body* which is either an action or a wff and finally, an optional *Gate* which is a wff. The understanding here is that summaries are associated with agent and for an agent "x" to have summary "u", then there are three cases depending on the body of "u":

1. If the *Body* of "u" is a wff, then

   $\vdash$ **(BEL x (ALWAYS** *(Gate* $\wedge$ *Body)* $\supset$ *(Gate* $\wedge$ *Effect)*)**)** [15]

2. If the *Body* of "u" is an action term, then

   $\vdash$ **(BEL x (***Gate* $\supset$ **(RESULT agent** *Body (Gate* $\wedge$ *Effect)*)**)))**

[15] Of course, many actions change the truth of their preconditions. Handling such actions and preconditions is straightforward.

```
1.    (DONE [(MK x y (ATTEND y x)) ∧
          (•conditions)]?;
      [IMPER x y "do y act"])                             Given
2.    (BMB y x (GOAL x (BEL y (GOAL x
                  (P-GOAL y (DONE y act)))))) ∧           P27, P3, P4, 1
          •(BMB y x (SINCERE x
                (GOAL x (P-GOAL y (DONE y act)))))
3.    (BMB y x (GOAL x (P-GOAL y (DONE y act)))) ∧        P11, P12, 2
          •(BMB y x (ALWAYS
                (COMPETENT y (DONE y act))))
4.    (BMB y x (GOAL x ◇[(DONE y act) ∨
                (BEL y (ALWAYS ~(DONE y act)))])) ∧       T1, P16, 3
          •(BMB y x ~(ALWAYS ~(DONE y act)))
5.    (BMB y x (GOAL x ◇(DONE y act))) ∧                  P16, P20, P8, 4
6.    (BEL y x (GOAL x ◇(DONE y act))) ∧                  Def. BMB
                (HELPFUL y x)
7.    (P-GOAL y x (DONE y act)) ∧                         Def. of HELPFUL, MP
          •(ALWAYS (COMPETENT y (DONE y act)))
8.    ◇[(DONE y act) ∨ (BEL y (ALWAYS ~(DONE y act)))] ∧  T1
          • ~(ALWAYS ~(DONE y act))
9.    ◇(DONE y act)                                       P20, P8
Q.E.D.
```

Figure 1: *Proof of Theorem 2 — An imperative to do an act results in its eventually being done.* [14]

One thing worth noting about summaries is that normally the wffs used above

⊢ (BEL x (*Gate* ⊃ ...))

will follow from the more general wff

⊢ *Gate* ⊃ ...

However, this need not be the case and different agents could have different summaries (even with the same name). Saying that an agent has a summary is no more than a convenient way of saying that the agent always believes an implication of a certain kind.

7 Summarization of a Request

The following is a summary named **REQUEST** that captures steps 2 through steps 5 of the proof of Theorem 2.

```
[REQUEST x y act]:
Gate:   (1) (BMB y x (SINCERE x (GOAL x
                (P-GOAL y (DONE y act))))) ∧
        (2) (BMB y x (ALWAYS
                (COMPETENT y (DONE y act))))
        (3) (BMB y x ~(ALWAYS ~(DONE y act)))
Body:     (BMB y x
            (GOAL x
              (BEL y
                (GOAL x (P-GOAL y (DONE y act))))))
Effect:   (BMB y x (GOAL x ◇(DONE y act)))
```

This summary allows us to conclude that any action preserving the *Gate* and making the *Body* true makes the *Effect* true. Conditions (2) and (3) are theorems and hence are always preserved. Condition (1) was preserved by assumption.

Searle's conditions for requesting are captured by the above. Specifically, his "propositional content" condition, which states that one requests a future act, is present as the *Effect* because of Theorem 2. Searle's first "preparatory" condition — that the hearer be able to do the requested act, and that the speaker think so is satisfied by condition (2). Searle's second preparatory condition — that it not be obvious that the hearer was going to do the act anyway — is captured by our conditions on persistence, which state when an agent can give up a persistent goal, that is not one of maintenance, when it has been satisfied.

Grice's "recognition of intent" condition [12, 13] is satisfied since the endpoint in the chain (step 9) is a goal. Hence, the speaker's goal is to get the hearer to do the act by means, in part, of the (mutual) recognition that the speaker's goal is to get the hearer to do it. Thus, according to Grice, the speaker has *meant*$_{nn}$ that the hearer should do the act. Searle's revised Gricean condition, that the hearer should "understand" the literal meaning of the utterance, and what illocutionary act the utterance "counts as" are also satisfied, provided the summary is mutually known. [16]

7.1 Nonserious Requests

Two questions now arise. First, is this not overly complicated? The answer, perhaps surprisingly, is "No". By applying this **REQUEST** theorem, we can prove that the utterance of an imperative in the circumstances specified by the *Gate* results in the *Effect*, which is as simple a propositional attitude as anyone would propose for the effect of uttering an imperative — namely that it is mutually believed that the speaker's goal is that the hearer eventually do the act. The *Body* need never be considered

[16] The further elaboration of this point that it deserves is outside the scope of this paper.

unless one of the gating conditions fails.

Then, if the *Body* is rarely needed, when is the "extra" embedding (**GOAL speaker (BEL hearer** ...)) attitude of use? The answer is that these embeddings are essential to preventing nonserious or insincere imperatives from being interpreted *unconditionally* as requests. In demonstrating this, we will show how Searle's "Sincerity" condition is captured by our **SINCERE** predicate.

The formula (**SINCERE speaker p**) is false when the speaker does something to get the hearer to believe he, the speaker, has the goal of the hearer's believing **p**, when he in fact does not have the goal of the hearer's knowing that **p** Let us see see how this would be applied for "Go jump in the lake", uttered idiomatically. Notice that it could be uttered and meant as a request, and we should be able to capture the distinction between serious and nonserious uses. In the case of uttering this imperative, the content of **SINCERE**, **p p** = (**GOAL speaker (P-GOAL hearer (DONE hearer [JUMP-INTO Lake1])))**.

Assume that it is mutually known/believed that the lake is frigidly cold (any other conditions leading to ~(**GOAL x p**) would do as well, e.g., that the hearer is wearing his best suit, or that there is no lake around). So, by a reasonable axiom of goal formation, no one has goals to achieve states of affairs that are objectionable (assume what is "objectionable" involves a weighing of alternatives). So, it is mutually known/believed that ~(**GOAL speaker (DONE hearer [JUMP-INTO Lake1]))**, and so the speaker does not believe he has such a goal. [17] The consequent to the implication defining **SINCERE** is false, and because the result of the imperative is a mutual belief that the speaker's goal is that the hearer think he has the goal of the hearer's jumping into the lake, the antecedent of the implication is true. Hence, the speaker is insincere or not serious, and a request interpretation is blocked. [18]

In the case of there not being a lake around, the speaker's goal cannot be that the hearer form the persistent goal of jumping in some non-existent lake, since by the *No Futility* property, the hearer will not adopt a goal if it is unachievable, and hence the speaker will not form his goal to achieve the unachievable state of affairs (that the hearer adopt a goal he cannot achieve). Hence, since all this is mutually believed, using the same argument, the speaker must be insincere.

8 Nonspecific requests

The ability conditions for requests are particularly simple, since as long as the hearer knows what action the speaker is referring to, he can always do it. He cannot, however, always bring about some goal world. An important variation of requesting is one in which the speaker does not specify the act to be performed; he merely expresses his goal that some **p** be made true. This will be captured by the action [**IMPER y 'p**] for "make **p** true". Here, in planning this act, the speaker need only believe the hearer thinks it is mutually believed that it is always the case that the hearer will eventually find a plan to bring about **p**. Although we cannot present the proof that performing an [**IMPER x y 'p**] will make $\Diamond$**p** true, the following is the illocutionary summary of that proof: [**NONSPECIFIC-REQUEST x y p**]:

Gate: (**BMB y x (SINCERE x (GOAL x (BEL y (GOAL x (P-GOAL y p))))))** $\wedge$
(**BMB y x (ALWAYS (COMPETENT y p)))**
(**BMB y x (ALWAYS** $\Diamond\exists$ **act' (DONE y q?;act'),**
where **q** $\stackrel{def}{=}$ (**BEL y (RESULT y act' p))))**

Body: (**BMB y x (GOAL x (BEL y (GOAL x (P-GOAL y p)))))**

Effect: (**BMB y x (GOAL x** $\Diamond$**p))**

Since the speaker only asks the hearer to make **p** true, the ability conditions are that the hearer think it is mutually believed that it is always true that eventually there will be some act such that the hearer believes of it that it achieves **p** (or he will believe it is impossible for him to achieve). The speaker need not know what act the hearer might choose.

9 On summarization

Just as mathematicians have the leeway to decide which proofs are useful enough to be named as lemmas or theorems, so too does the language user, linguist, computer system, and speech act theoretician have great leeway in deciding which summaries to name and form. Grounds for making such decisions range from the existence of illocutionary verbs in a particular language, to efficiency. However, summaries are flexible — they allow for different languages and different agents to carve up the same plans differently. [19] Furthermore, a summary formed for efficiency may not correspond to a verb in the language.

Philosophical considerations may enter into how much of a plan to summarize for an illocutionary verb. For example, most illocutionary acts are considered successful when the speaker has communicated his intentions, not when the intended effect has taken hold. This argues for labelling as *Effects* of summaries intended to capture illocutionary acts only formulas that are of the form (**BMB hearer speaker (GOAL speaker p))**, rather than those of the form (**BMB hearer speaker p**) or (**BEL hearer p**), where **p** is not a **GOAL**-dominated formula. Finally, summaries may be formed as conversations progress.

The same ability to capture varying amounts of a chain of inference will allow us to deal with multi-utterance or multi-agent acts, such as, betting, complying, answering, etc., in which there either needs to be more than one act (a successful bet requires an offer and an acceptance), or one act is defined to require the presence of another (complying makes sense only in the presence of a previous directive). For example, where **REQUEST** captured the chain of inference from step 2 to step 5, one called **COMPLY** could start at 5 and stop at step 9.

[17] The speaker's expressed goal is that the hearer form a persistent goal to jump in the lake. But, by the *Inevitable Consequences* lemma, given that a c.o.e. satisfying the speaker's goal also has the hearer's eventually jumping in (since the hearer knows what to do), the speaker's goal is also a c.o.e. in which the hearer eventually jumps in. In the same way, the speaker's goal would also be that the hearer eventually gets wet.

[18] However, we do not say what else might be derivable. The speaker's true goals may have more to do with the manner of his action (e.g., tone of voice), than with the content. All we have done is demonstrate formally how a hearer could determine the utterance is not to be taken at face value.

[19] Remember, summaries are actually beliefs of agents, and those beliefs need not be shared.

Thus, the notion of characterizing illocutionary acts as lemma-like summaries, i.e., as chains of inference subject to certain conditions, buys us the ability to encapsulate distant inferences at "one-shot".

9.1 Ramifications for Computational Models of Language Use

The use of these summaries provides a way to prove that various short-cuts that a system might take in deriving a speaker's goals are correct. Furthermore, the ability to index summaries by their *Bodies* or from the utterance types that could lead to their application (e.g., for utterances of the form "Can you do <X>") allows for fast retrieval of a lemma that is likely to result in goal recognition. By an appropriate organization of summaries [5], a system can attempt to apply the most comprehensive summaries first, and if inapplicable, can fall back on less comprehensive ones, eventually relying on first principles of reasoning about actions. Thus, the apparent difficulty of reasoning about speaker-intent can be tamed for the "short-circuited" cases, but more general-purpose reasoning can deployed when necessary. However, the complexities of reasoning about others' beliefs and goals remains.

10 Extensions: Indirection

Indirection will be modeled in this framework as the derivation of propositions dealing with the speaker's goals that are not stated as such by the initial propositional attitude. For example, if we can conclude from **(BMB y x (GOAL x (GOAL y p)))** that **(BMB y x (GOAL x (GOAL y $\Diamond$ q)))**, where p does not entail q, then, "loosely", we will say an indirect request has been made by x.

Given the properties of $\Diamond$, **(GOAL x p) $\supset$ (GOAL x $\Diamond$p)** is a theorem. **(GOAL x p)** and **(GOAL x ~p)** are mutually unsatisfiable, but **(GOAL x $\Diamond$p)** and **(GOAL x $\Diamond$~p)** are jointly satisfiable. For example, **(GOAL BILL $\Diamond$HAVE BILL HAMMER1)))** and **(GOAL BILL $\Diamond$(HAVE JOHN HAMMER1))** could both be part of a description of Bill's plan for John to get a hammer and give it to him. Such a plan could be triggered by Bill's merely saying "Get the hammer" in the right circumstances, such as when Bill is on a ladder plainly holding a nail. [20] A subsequent paper will demonstrate the conditions under which such reasoning is sound.

11 Concluding Remarks

This paper has demonstrated that all illocutionary acts need not be primitive. At least some can be derived from more basic principles of rational action, and an account of the propositional attitudes affected by the uttering of sentences with declarative, interrogative, and imperative moods. This account satisfies a number of criteria for a good theory of illocutionary acts.

- Most elements of the theory are independently motivated. The theory of rational action is motivated independently from any notions of communication. Similarly, the properties of cooperative agents are also independent of communication.

[20] Notice that most theories of speech acts would treat the above utterance as *only* a direct request. We do not.

- The characterization of the result of uttering sentences with certain syntactic moods is justified by the results we derive for illocutionary acts, as well as the results we cannot derive (e.g., we cannot derive a request under conditions of insincerity).

- Summaries need not correspond to illocutionary verbs in a language. Different languages could capture different parts of the same chain of reasoning, and an agent might have formed a summary for purposes of efficiency, but that summary need not correspond to any other agent's summary.

- The rules of combination of illocutionary acts (characterizing, for example, how multiple assertions could constitute the performance of a request) are now reduced to rules for combining propositional contents and attitudes. Thus, multi-utterance illocutionary acts can be handled by accumulating the speaker's goals expressed in multiple utterances, to allow an illocutionary theorem to be applied.

- Multi-act utterances are also a natural outgrowth of this approach. There is no reason why one cannot apply multiple illocutionary summaries to the result of uttering a sentence. Those summaries, however, need not correspond to illocutionary verbs.

- The theory is naturally extensible to indirection (to be argued for in another paper), to other illocutionary act, such as questions, commands, informs, assertions, and to the act of referring [8].

Finally, although illocutionary act recognition may be strictly unnecessary, given the complexity of our proofs, it is likely to be useful. Essentially, such recognition would amount to the application of illocutionary summaries theorems to discover the speaker's goal(s).

12 Acknowledgements

We would like to thank Tom Blenko, Herb Clark, Michael Georgeff, David Israel, Bob Moore, Geoff Nunberg, Fernando Pereira, Ray Perrault, Stan Rosenschein, Ivan Sag, and Moshe Vardi for valuable discussions.

13 References

1. Allen, J. F. A plan-based approach to speech act recognition. Technical Report 131, Department of Computer Science, University of Toronto, January, 1979.

2. Allen, J. F., Frisch, A. M., & Litman, D. J. ARGOT: The Rochester dialogue system. *Proceedings of the National Conference on Artificial Intelligence*, Pittsburgh, Pennsylvania, 1982, 66-70.

3. Appelt, D. *Planning Natural Language Utterances to Satisfy Multiple Goals.* Ph.D. Th., Stanford University, Stanford, California, December 1981.

4. Austin, J. L. *How to do things with words.* Oxford University Press, London, 1962.

5. Brachman, R., Bobrow, R., Cohen, P., Klovstad, J., Webber, B. L., & Woods, W. A. Research in natural language understanding. Technical Report 4274, Bolt Beranek and Newman Inc., August, 1979.

6. Bruce, B. C., & Newman, D. Interacting plans. *Cognitive Science 2*, 3, 1978, pp. 195-233.

7. Clark, H. H., & Marshall, C. Definite reference and mutual knowledge. In *Elements of Discourse Understanding*, Academic Press, Joshi, A. K., Sag, I. A., & Webber, B., Eds., New York, 1981.

8. Cohen, P. R. The Pragmatics of Referring and the Modality of Communication. *Computational Linguistics 10*, 2, 1984, pp. 97-146.

9. Cohen, P. R. *On Knowing what to Say: Planning Speech Acts.* Ph.D. Th., University of Toronto, Toronto, January 1978. Technical Report No. 118, Dept. of Computer Science.

10. Cohen, P. R., & Levesque, H. J. Speech Acts and the Recognition of Shared Plans. *Proc. of the Third Biennial Conference*, Canadian Society for Computational Studies of Intelligence, Victoria, B. C., May, 1980, 263-271.

11. Emerson, E. A., and Halpern, J. Y. "Sometimes" and "Not Never" Revisited: On Branching versus Linear Time. *ACM Symposium on Principles of Programming Languages*, 1983.

12. Grice, H. P. Meaning. *Philosophical Review 66*, 1957, pp. 377-388.

13. Grice, H. P. Utterer's Meaning and Intentions. *Philosophical Review 68*, 2, 1969, pp. 147-177.

14. Grice, H. P. Logic and conversation. In Cole., P. and Morgan, J. L., Eds.,*Syntax and Semantics: Speech Acts* , Academic Press, New York,1975.

15. Halpern, J. Y., and Moses, Y. O. A Guide to the Modal Logics of Knowledge and Belief. *Proc. of the Ninth International Joint Conference on Artificial Intelligence*, IJCAI, Los Angeles, Calif., August, 1985.

Levesque, Hector, J. A logic of implicit and explicit belief. *Proceedings of the National Conference of the American Association for Artificial Intelligence*, Austin, Texas, 1984.

esque, H. J., & Cohen, P. R. A Simplified Logic of Interaction. In preparation

18. McCarthy, J., & Hayes, P. J. Some Philosophical Problems from the Standpoint of Artificial Intelligence. In *Machine Intelligence 4*, American Elsevier, B. Meltzer & D. Michie, Eds., New York, 1969.

19. McDermott, D. A temporal logic for reasoning about processes and plans. *Cognitive Science 6*, 2, 1982, pp. 101-155.

20. Moore, R. C. Reasoning about Knowledge and Action. Technical Note 191, Artificial Intelligence Center, SRI International, October, 1980.

21. Morgan, J. L. Two types of convention in indirect speech acts. In *Syntax and Semantics, Volume 9: Pragmatics*, Academic Press, P. Cole, Ed., New York, 1978, 261-280.

22. Perrault, C. R., & Allen, J. F. A Plan-Based Analysis of Indirect Speech Acts. *American Journal of Computational Linguistics 6*, 3, 1980, pp. 167-182.

23. Perrault, C. R., & Cohen, P. R. It's for your own good: A note on inaccurate reference. In *Elements of Discourse Understanding*, Cambridge University Press, Joshi, A., Sag, I., & Webber, B., Eds., Cambridge, Mass., 1981.

24. Schiffer, S. *Meaning*. Oxford University Press, London, 1972.

25. Schmidt, D. F., Sridharan, N.S., & Goodson, J. L. The plan recognition problem: An intersection of artificial intelligence and psychology. *Artificial Intelligence 10*, 1979, pp. 45-83.

26. Searle, J. R. *Speech acts: An essay in the philosophy of language*. Cambridge University Press, Cambridge, 1969.

27. Sidner, C. L., Bates, M., Bobrow, R. J., Brachman, R. J., Cohen, P. R., Israel, D. J., Webber, B. L., & Woods, W. A. Research in knowledge representation for natural language understanding. Annual Report 4785, Bolt, Beranek and Newman Inc., November, 1981.

28. Vanderveken, D. A Model-Theoretic Semantics for Illocutionary Force. *Logique et Analyse 26*, 103-104, 1983, pp. 359-395.

13 Appendix

Proof of Theorem 1:

First, we need a lemma:

Lemma 3 ∀a (DONE x [(BEL x (AFTER a p)) ∧ (COMPETENT x p)]?;a) ⊃ p

Proof:

1.	∀a (DONE x [(BEL x (AFTER a p)) ∧ (COMPETENT x p)]?;a)	Ass
2.	(BEL x (AFTER x p)) ∧ (COMPETENT x p) ⊃ (AFTER a p)	Def. of COMPETENT, MP
3.	∀a (DONE x (AFTER a p)?;a)	2, P4
4.	p	3, P3
5.	∀a (DONE x [(BEL x (AFTER a p)) ∧ (COMPETENT x p)]?;a) ⊃ p	Impl. Intr.

Q.E.D.

Theorem 1. ∀y (P-GOAL y p) ∧ (ALWAYS (COMPETENT y p)) ⊃ ◇(p ∨ (BEL y (ALWAYS ~p)))

Proof:

1.	(P-GOAL y (DONE y act)) ∧ (ALWAYS (COMPETENT y (DONE y act)))	Ass.
2.	◇(∃a (DONE y [(BEL y (AFTER a p))]?;a) ∨ (BEL y (ALWAYS ~p)))	1, P25, MP
3.	◇(p ∨ (BEL y (ALWAYS ~p)))	L3, P8, 2
4.	(P-GOAL y (DONE y act)) ∧ (ALWAYS (COMPETENT y (DONE y act))) ⊃ ◇(p ∨ (BEL y (ALWAYS ~p)))	Impl. Intr., 3

Q.E.D.