Rational Interaction as the Basis for Communication*

Philip R. Cohen
Artificial Intelligence Center
and
Center for the Study of Language and Information
SRI International
and
Hector J. Levesque†
Department of Computer Science
University of Toronto

March 14, 1988

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 neither be primitive, nor explicitly recognized. As a test case, we derive Searle's conditions on requesting from principles of rationality coupled with a theory of imperatives. The theory rests on a formal account of intention and distinguishes insincere or nonserious imperatives from true requests. A theory of purposeful communication thus emerges as a consequence of principles of action and interaction.

*This research was made possible by a gift from the Systems Development Foundation, by a grant from the Natural Sciences and Engineering Research Council of Canada, by the Defense Advanced Research Projects Agency under Contract N00039-84-K-0078 with the Naval Electronic Systems Command, and by a contract from the Nippon Telegraph and Telephone Corp. The paper is approved for public release, distribution unlimited. 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, the United States government, the Canadian government, or NTT Corporation. This paper will appear in *Intentions in Communication*, Cohen, P. R., Morgan, J., and Pollack, M. E., eds., MIT Press, 1988.

†Fellow of the Canadian Institute for Advanced Research.

2 Introduction

This paper explores the consequences of viewing language as action. This approach provides us with not just a slogan, but rather a program of research directed at identifying those aspects of language use that follow from general principles of rational, cooperative interaction. Our pursuit of such a program does not mean that we believe all language use is completely and consciously thought out and planned. Far from it. Rather, just as there are grammatical, processing, and sociocultural constraints on language use, so may there be constraints imposed by the rational balance that agents maintain among their beliefs, intentions, commitments, and actions. Our goals are to discover such constraints, to develop a logical theory that incorporates them and predicts dialogue phenomena, and finally to apply them in developing algorithms for human-computer interaction in natural language.

In our pursuit of this research, we treat utterance as instances of other events that change the state of the world; utterance events per se change the mental states of speakers and hearers. Utterance events are typically performed by a speaker to effect such changes. Moreover, they do so because they signal, or convey (at least) the information that the speaker is in a certain mental state, such as intending the hearer to adopt a certain mental state. Conversations are initiated and proceed because of an interplay among agents' mental states, their capabilities for purposeful behavior, their cooperativeness, the content and circumstances of their utterances, and other factors that surely remain to be elucidated. A theory of conversation based on this approach would explain dialogue coherence in terms of the participants' mental states, how the latter lead to communicative action, how these acts affect the mental states of the hearers, and so on.

2.1 Thesis: Illocutionary Force Recognition is Unnecesary

Speech act theory appears to offer a natural route leading toward some of these goals. After all, it is in this context that theorists have promoted, and in some depth examined, many of the implications of treating language as action. Speech act theory was originally conceived as part of action theory. Many of Austin's insights about the nature of speech acts, felicity conditions, and modes of failure were derived from a study of noncommunicative actions. Searle [41] mentions repeatedly that many of the conditions he attributes to various illocutionary acts (such as requests and questions) apply more generally to noncommunicative action. However, in recent work Searle and Vanderveken [42] (hereafter S&V) formalize communicative acts and propose a logic in which their properties, such as "preparatory conditions" and "modes of achievement," are stipulated primitively, rather than being derived from more basic principles of action. We believe such an approach overlooks significant generalities. Moreover, it leads one to build logics of illocutionary acts independently of theories of action. Our research shows how to derive properties of illocutionary acts from principles of rationality; hence it suggests that the theory of illocutionary acts is not explanatory, but descriptive.

Consider the following seemingly trivial dialogue fragment:

A: "Open the door."

B: "Sure"

From a syntactic standpoint, these utterances are uninteresting. Of course, the semantics and effects of imperatives (which we shall explain) are nontrivial, while the meaning of "Sure" is unclear. Yet it seems that the speakers' intentions and the situations in which their utterances are made play the crucial role in determining what has happened during the dialogue and how what has changed can influence agents' subsequent actions. It would be reasonable to *describe* what has happened by saying that A has performed a directive speech act (e.g., a request) and that B has performed a commissive (e.g., a promise). To verify that B did in fact do this, imagine B's saying "Sure" and then doing nothing. A would surely be justified in complaining or asking for an explanation. A competence theory of communication needs to elucidate just how an interpersonal commitment becomes established. The theory presented in this paper does so by explaining what effects are brought about by a speaker's uttering an imperative in a given situation, and how the uttering of "Sure" relates to those effects. These explications will make crucial reference to intention, but need not involve the hearer's recognizing which illocutionary acts were performed.

It is tempting to read (or perhaps misread), philosophers of language as saying that illocutionary-force recognition is necessary for successful communication. Austin [4] and Strawson [43] require that "uptake" take place. Searle and Vanderveken [41,42] contend that illocutionary force is part of the meaning of an utterance and that the latter's intended effect is "understanding." Hence, because hearers are intended to understand the utterance, presumably its meaning, one interpretation of Searle and Vanderveken's claim is that the hearer is intended to recognize the utterance's illocutionary force.[1] Bach and Harnish [5] also make a similar claim.[2]

It is so tempting to read these writers thus that many researchers, including us, have made this assumption. For example, computational models of dialogue [1,2,7] that we and our colleagues have developed have required the computer program to recognize which illocutionary act the user has performed so that the system can respond as intended. However, we now claim that force recognition is usually unnecessary. For example, in both of the systems mentioned above, all the inferential power of the recognition of illocutionary acts was already available from other inferential sources [17]. Instead, we argue that many properties of illocutionary acts can be *derived* from the speaker's and hearer's mental states, especially from his beliefs and intentions. What speakers and hearers have to do is only to recognize

[1] But perhaps they mean illocutionary-force *potential*. They write: "Part of the meaning of an elementary sentence is that its literal utterance in a given context constitutes the performance or attempted performance of an illocutionary act of a particular illocutionary force." [42, p.7]. The question at issue here is whether, as a hearer's understands an utterance and knows its meaning, he recognizes (or is intended to recognize) that the specific utterance in that specific context was uttered with a specific illocutionary force. The following remark leads us to believe the answer is affirmative: "But, I wish to claim, the intended effect of meaning something is that the hearer should know the illocutionary force and propositional content of the utterance ..." [40, p. 8].

[2] "What sort of explanation does the hearer seek of the speaker's utterance? ... However, he seeks also to identify the locutionary and illocutionary act performed by the speaker in his utterance, and this involves ascribing intentions to the speaker, in particular, the intention to be performing a certain illocutionary act (by way of performing a certain locutionary act)." [5, p.89].

each other's intentions (based on mutual beliefs). Contrary to other proposed theories, we do not require that those intentions include intentions that the hearer recognize precisely which illocutionary act(s) were being performed.

Although one can *label* parts of a discourse with names of illocutionary acts, illocutionary labeling does not constitute an explanation of a dialogue. Rather, the labeling itself, if reliably obtained, constitutes data to be explained by constraints on mental states and actions. That is, one would show how to derive the labelings, given their definitions, from (for example) the beliefs and intentions the participants are predicted to have by the analysis of the preceding interaction. Although hearers *may* find it heuristically useful to determine just which illocutionary act was performed, our view is that illocutionary labeling is an extra task in which dialogue participants may be able to engage only retrospectively.

The view that illocutionary acts are not primitive and there need not be recognized explicitly is a liberating one. Once this position is adopted, it becomes apparent that many of the difficulties in applying speech act theory to discourse or incorporating it into computer systems, stem from taking these acts too seriously — i.e., as primitives.

2.2 Illocutionary Actions as Complex Event-Types

Despite Austin's concern for speakers' performance of illocutionary acts by means of locutionary acts, most of the interesting speech act theories have dealt primarily with the illocutionary act. In so doing, theorists have treated (perhaps out of convenience) illocutionary acts as unitary and nondecomposable primitives, though subject to many conditions. For example, Searle's [41] analysis provided necessary and sufficient conditions for the nondefective and successful performance of illocutionary acts. Early linguistic analyses attempted to derive illocutionary classifications via transformations of implicit performative elements [37] or by conversational postulates applied to primitive illocutionary act elements [21].

One sees this view of illocutionary acts as primitives most clearly in examining various treatments of indirect speech acts. In classifying the utterance, "Can you reach the hammer?" in terms of illocutionary acts, the speaker's questioning the hearer's ability to reach the hammer and his requesting that the hearer pass the hammer are regarded as different actions he may be performing simultaneously. For Searle [38] and Bach and Harnish [5], an analysis of indirect speech acts is concerned with specifying how, for example, such a request can be made by means of the illocutionary act of questioning. That is, the nub of the analysis rests on uncovering relationships amongst illocutionary acts. Other difficult problems for a theory of speech acts that arise when the primary unit of analysis is the illocutionary act include specifying how multiple illocutionary acts not in a by-means-of relation can occur simultaneously, and how multiple utterance acts can somehow constitute the performance of one illocutionary act. To address these problems, one needs a calculus of acts.[3]

Goldman [20] attempts to provide such a calculus by giving an inductive definition of the "generation" relation that holds among actions, roughly, where one action can be said to be

[3]The words "action" and "act" are sometimes used in philosophical writings to make a type/token distinction. We shall not adopt this usage, preferring to let context, and ultimately the formalism, disambiguate the intended meaning.

done "by" doing another. Based on a notion of primitive action, Goldman inductively defines actions that are generated by those primitives using four types of generation relationships — *causal*, *conventional*, *simple*, and *augmentation* generation. It would take us too far afield to provide the definitions (but see [34,35] for further discussion). Goldman takes the view that agents perform indefinitely many actions when they do anything. Consider Searle's [39] examples of Gavrilo Princip's pulling a trigger, firing a gun, killing Archduke Ferdinand, and starting World War I. According to Goldman, all the actions that Princip does are different. Goldman contrasts this approach with that of Davidson [19], who argues that agents perform specific events, about which one can have many different descriptions (which can be regarded as terms in a formal language). Thus, Princip does one thing, pull the trigger, and his firing the gun, his killing the archduke, and his starting World War I are all different descriptions of that event. Goldman finds problems with this approach since descriptions denoting the same entity (the event) should be intersubstitutable yet preserve truth. This clearly doesn't hold; pulling a trigger causes a bullet to be emitted, but killing the Archduke does not.

Clearly, there is something to be said for both approaches — for Goldman's use of complex properties to describe actions, and for Davidson's intuition that only one act was performed. We advocate a position incorporating some of the advantages of each approach, namely that agents perform single instances of primitive act- or event-types (ignoring simultaneous events), but each specific act or event can realize many different complex actions. We have chosen to develop a characterization of complex event-types based on operators familiar from the computer science literature, namely temporal and dynamic logics. The chosen operators are not wholly satisfactory (e.g., they do not characterize simultaneous action), but they have a precise semantics and can approximate Goldman's relations adequately for our purposes.

Essentially, we assume a set of primitive event-types, which incorporate the agent and other intrinsic arguments, but are abstracted over time. So, an instance of a primitive event-type can occur more than once. Then, we allow the formation of complex event-types, characterized by *action expressions* in the formal language we develop, in terms of composition via sequence, disjunction, and circumstance. The latter allows us to describe the situation-specific effects that result when events occur in certain contexts. Moreover, the treatment of action is sufficient to model conditional and iterative actions, an important desiderata for any theory of action. Thus, our complex event types describe sequences of instances of primitive event-types occurring in various circumstances.

On the basis of such a logic of action, one ought to be able to derive the properties of the complex event-types from the properties of their defining elements. In the domain of illocutionary acts, given an utterance event performed in the context of the speaker's and hearer's specific mental states, the theorist ought to be able to determine which illocutionary actions were performed and what relationships exist among them from an analysis of the situation-specific effects of an utterance event. So, for example, we would explain indirection by showing how the direct and indirect illocutionary act classifications are both derivable from the utterance event, given the circumstances and properties of rational interaction. We would not attempt to derive the indirect illocutionary act classification from the direct one. The purpose of this paper is to show how properties of illocutionary acts can be derived from

a sufficiently detailed logic of action.

3 Form of the Argument

We demonstrate the fact that illocutionary actions can be treated as action expressions by deriving Searle's conditions on illocutionary acts from an independently motivated theory of action. The realm of communicative action is entered in accordance with Grice's method [24]: by postulating a correlation between the uttering of a sentence with a certain syntactic feature (e.g., with its dominant clause an imperative) in a certain context, and a complex propositional attitude expressing the speaker's mental state. As a result of the speaker's uttering a sentence with that feature under those conditions, the hearer comes to have various beliefs (assumptions) that the speaker has the corresponding attitude. Because of general principles governing mental states, other consequences of the speaker's having the expressed state can be derived. Such derivations will be used to form complex action expressions that capture illocutionary acts in terms of the speaker's attempting to bring about some part of the chain of consequences by bringing about an antecedent. For example, the action expression to be called REQUEST will encapsulate a derivation in which a speaker attempts to have (1) the hearer form the intention to act because (2) it is mutually believed the speaker wants him to act. The conditions justifying the inference from (2) to (1) can be shown to subsume those claimed by Searle [41] to be felicity conditions. However, they have been derived here from first principles and without any need for a primitive action of requesting. Moreover, they satisfy a set of adequacy criteria, which consist of the following: differentiating the form of an utterance from its illocutionary force; handling the major kinds of illocutionary acts; modeling a speaker's insincere performance of illocutionary acts; providing an analysis of performative utterances, showing how illocutionary acts can be performed with multiple utterances, how multiple illocutionary acts can be simultaneously performed with one utterance, and explaining indirect speech acts.

Our approach is similar to that of Bach and Harnish [5] in its reliance on inference. A theory of rational interaction will provide the formal foundation for drawing the necessary inferences. A notion of sincerity is essential for treating deception and nonserious utterances. Finally, a characterization of utterance features (e.g., mood) is required in making a transition from the domain of utterance syntax and semantics, to that of utterance effects (on speakers and hearers). There are three main steps in constructing the theory:

1. *Infer illocutionary point from utterance form.* The theorist derives the chains of inference needed to connect the intentions and beliefs signaled by an utterance's form with typical "illocutionary points" [42], such as getting a hearer to do some action. These derivations are based on principles of rational interaction and are independent of theories of speech acts and communication.

Specifically (referring to Figure 1), assume that actions (A) are characterized as producing certain effects E_1 when executed in circumstances C. Separately, assume that the theorist has either derived or postulated relationships between effects of type E_{i-1} and other effects, say of type E_i such that, if E_{i-1} holds in the presence of some gating condition C_{i-1}, then

$$C : A \rightarrow E_1 \stackrel{C_1}{\rightarrow} E_2 \stackrel{C_2}{\rightarrow} E_3 \rightarrow \ldots \stackrel{C_{i-1}}{\rightarrow} E_i$$

Figure 1: Events producing gated effects

E_i holds as well. One can then prove that, in the right circumstances — specifically those satisfying the gating conditions — doing action A makes E_i true.[4]

2. *Treat illocutionary acts as attempts.* Searle [41] points out that many communicative acts are attempts to achieve some effect. For example, requests are attempts to get (in a certain way) the hearer to do some action. Roughly, we shall say that an agent *attempts* to achieve some state of affairs E_i if he performs some action or sequence of actions A that, he intends, will bring about effect E_i. The intended effect may not be an immediate consequence of the utterance act, but could be related to act A by some chain of causally related effects. Under these conditions, for A to be an attempt to bring about E_i, the agent would have to want the gating conditions C_i to hold after A and that consequently, in doing A in circumstances C, he wants E_i to obtain.

3. *Create an action expression to capture the illocutionary-act type.* This expression will involve events performed in the context of a speaker's attempting to achieve various effects. Because illocutionary acts can be performed through utterances of different forms, we abstract from any specific type of utterance event in defining illocutionary acts.

This way of treating communicative acts has many advantages. The framework clarifies the degrees of freedom available to the theorist by showing which properties of communicative acts are consequences of independently motivated elements and which properties are stipulated. Furthermore, it shows the freedom available to linguistic communities in naming patterns of inference as illocutionary verbs. It also lends technical substance to the use of such terms as "counts as, " "felicity conditions," and "illocutionary force." However, it makes no commitment to a reasoning strategy. For example, the theorist's derivations from first principles may be encapsulated by speakers and hearers as frequently used lemmas. Speakers and hearers need not in fact believe that the gating conditions hold, but may instead assume that they hold and then "jump" to the consequent of the lemma. The key to achieving these goals is to have an adequate analysis of intention, one that relates intending to other mental states as well as to the agent's actions. We sketch our theory of intention below; further expressions can be found in Chapter 2 and in other publications of ours [15].

We model intention as a composite concept specifying what an agent has *chosen* and how he is *committed* to that choice. First, consider the case of an agent choosing from his [possibly inconsistent] desires those he wants most to see fulfilled. In a loose sense, let us call these chosen desires goals.[5] By assumption, chosen desires are consistent. We will give them a possible-world semantics and the agent will thus have selected a set of worlds in which the

[4] Another way to characterize utterance effects is by applying "default logic" [31].

[5] Chosen desires are ones that speech act theorists claim to be conveyed by such illocutionary acts as requests.

goals hold.

Next, consider an agent to have a *persistent goal* if he has a goal (i.e., a proposition true in all of the agent's chosen worlds) that he believes currently to be false and that he will continue to choose — at least as long as certain facts remain valid. Persistence involves an agent's *internal* commitment over time to his choices.[6] For example, the complete fanatic is persistent until he believes his goal has been achieved or is impossible. The fanatical agent will drop his commitment to achieving the goal only if either of those circumstances holds.

Intention will be modeled as a kind of persistent goal — i.e., a persistent goal to do an action, believing one is about to do it, or to achieve some state of affairs, believing one is about to achieve it. When modeled this way, our concept of intention can be shown to (1) satisfy Bratman's [8,9,10] functional characteristics of intention and (2) to lack the undesirable trait of being closed under expected consequence [15].

Generally, intentions are formed against a background consisting, as a minimum, of agents' beliefs, desires, and other intentions. To capture this fact, we extend the concept of persistent goal and, by derivation, intention, so as to expand the conditions under which an agent can give up his goal. When necessary conditions for an agent's discarding of a goal include his having other goals (call them "supergoals"), the agent can generate a chain of goals such that, if the supergoals are given up, so may be the subgoals. If the conditions necessary for an agent's giving up a persistent goal include his believing that some *other* agent has a persistent goal, a chain of interpersonally linked goals is created. For example, if Mary asks Sam to do something and Sam agrees, Sam's goal should be persistent unless he finds out that Mary no longer wants him to do the requested action (or, as discussed earlier, he has done the action or has found it impossible). Both requests and promises are analyzed in terms of such "interpersonally relativized" persistent goals.

In summary, we provide an analysis of intention in terms of a concept of a persistent goal to perform an action. Intention is distinguished from the more atomic concept of "choice," modeled as chosen possible worlds. Choice is closed under expected consequence, whereas intention is closed only under logical equivalence and embodies a precise notion of commitment — if the agent fails to accomplish the intended action, the agent is committed to trying again (except under certain specified circumstances). Those readers interested in details of this formalism should read elsewhere [15]; the present exposition is *not* comprehensive.

In our subsequent presentation, we shall first give a brief synopsis of our theory of rational interaction, leading up to a discussion of the concepts of *persistent goal* and *intention*. We shall then characterize the effects of utterance events and, using persistent goals, define the notion of an attempt. Requests are then defined as attempts. To demonstrate the viability of the analysis, we show that it fulfills a substantive adequacy criterion: that it can be used to derive Searle's felicity conditions for requesting. Finally, we shall describe extensions of the formalism and theory that make it possible to handle other illocutionary acts.

[6] This is not a *social* commitment. It remains to be seen whether social commitments can be constructed from internal ones.

4 Elements of a Formal Theory of Rational Interaction

To achieve these goals, we need a carefully elaborated theory of rational action and interaction. The present section is a condensed version of the formalism presented by us elsewhere [15].

The analysis is expressed in a logic whose model theory is based on a possible-worlds semantics. We propose a logic with four primary modal operators — BEL[ief], GOAL, HAPPENS (a given action happens next) and DONE (a given action has just been performed). We shall use these operators to characterize what agents must know to perform actions that are intended to achieve their intentions. GOAL is introduced in order to model choice, intention, and commitment.

4.1 Syntax

The logic has the usual connectives of a first-order language with equality, as well as operators for propositional attitudes and for describing events that occur in various circumstances. Briefly, the following constitute the set of well-formed formulas:

- (BEL x p) and (GOAL x p), meaning that wff p follows from agent x's beliefs and goals, respectively;
- (HAPPENS a) and (DONE a), meaning that, at a given time, a event (or sequence of events) characterized by *action expression* a (see below) will happen next or has just happened, respectively. Notice that these formulas are true or false relative to a given time;
- (AGT x e), which is true if and only if agent x is the *only* agent of the sequence of events e;
- Time propositions, which are true if and only if the current time is the same as that expressed in the proposition;
- $e_1 \leq e_2$, which states that the sequence of events e_1 is an initial subsequence of the sequence of events e_2.

For convenience, let us define versions of DONE and HAPPENS that specify the agent of the act.

Definition 1 (DONE x a) $\stackrel{\text{def}}{=}$ (DONE a) $\wedge$ (AGT x a)

Definition 2 (HAPPENS x a)$\stackrel{\text{def}}{=}$ (HAPPENS a) $\wedge$ (AGT x a)

4.1.1 Events and Actions

The framework proposed here separates primitive from complex event-types. Examples of primitive event-types might include moving an arm, grasping, exerting a force, and uttering a word or sentence. Complex event-types are captured by *action expressions*, which are inductively defined from the primitives by the operators below. For example, the movement

of a finger may result in closure of a circuit, which may in turn result in a light's coming on. We shall say that one primitive event happened, an occurrence that can be characterized by various complex action expressions, or simply by *actions*.

Action Expressions: The following are action expressions, formed with the operators of dynamic logic [28,30,36]:

- Event variables e, e_1 e_2, ... e_n, ranging over sequences of primitive event-types.
- a;b, for sequential action composition, where a and b are action expressions.
- p?, the test action, where p is a wff. For example, the composition of the test action p? with an action expression a forms an action expression p?;a which describes a's being performed when the wff p is true.
- a|b, the nondeterministic choice action. This expression describes events that are described by either a or b.
- a^*, an iterative action.

With these operators, one can define the usual programming-language constructs for conditionals and while-loops.

Recall that (HAPPENS a) says that action a occurs next. When a occurs (next) in a context in which p holds, we write (HAPPENS p?;a).[7] To say that a brings about p (next), we use (HAPPENS ~p?;a;p?). This says that action a happens next and p is false, and right after a happens, p becomes true. Since there is no parallel activity in this model, this locution captures a's causing p to become true. To be a bit more concrete, one would normally not have a primitive-event type for closing a circuit. Therefore, to say John closed the circuit one would say that John did something (perhaps a sequence of primitive events) that caused the circuit to be closed — ∃x (DONE JOHN ~(CLOSED c)?;x;(CLOSED c)?).

Another way to characterize events is to express predications about them. For example, one could predicate (WALK e) to say that a given event-type e is a walking type of event. This way of describing events has the advantage of allowing complex properties, such as running a race, to hold for an undetermined (and unnamed) sequence of events. However, because the predications are made about the events, not the attendant circumstances, this method prevents us from describing events that happen only in certain circumstances. We shall have to employ both methods.

To describe the future, we define the usual linear temporal-logic operators, ◇ and □ based on HAPPENS:

Definition 3 *Eventually:* $\Diamond\alpha \stackrel{def}{=} \exists x$ (HAPPENS $x;\alpha?$).

[7] Notice that an indefinite number of other actions, created by the formation rules as applied to a have occurred too. For example, (HAPPENS (p∧p)?;a), etc.

In other words, $\Diamond\alpha$ is true (in a given possible world) if α holds after something that happens, that is, if α will be true at some point in the future.

Definition 4 *Always:* $\Box\alpha \stackrel{def}{=} \sim \Diamond \sim \alpha$.

$\Box\alpha$ means that henceforth α is true throughout the course of events. $\Box\alpha$ is the dual of $\Diamond\alpha$. Next, to talk about propositions that are not true now but will become true, we define the following:

Definition 5 (LATER p) $\stackrel{def}{=} \sim$p $\wedge \Diamond$p

We shall have occasion to state constraints on courses of events. To do so, we define the following:

Definition 6 (BEFORE p q) $\stackrel{def}{=} \forall$ c (HAPPENS c; q?) $\supset \exists$ a (a $\leq$ c) $\wedge$ (HAPPENS a; p?)

This definition states that p comes before q (starting at the current index n in the course of events) if, whenever q is true in the course of events, p has been true (after the index n).

Finally, one distinction should be pointed out. When action variables are bound by quantifiers, they range over sequences of events (more precisely, event types). When they are left free in a formula, they are schematic and can be instantiated with complex action expressions.

4.2 The Attitudes

BEL and GOAL characterize what is *implicit* in an agent's beliefs and goals (chosen desires), rather than what an agent believes actively or explicitly, or has as a goal.[8] That is, these operators characterize what *the world would be like* if the agent's beliefs and goals were true. It is important to note that we do not include an operator for wanting, since desires need not be consistent. Although desires certainly play a vital role in determining goals and intentions, 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.

For simplicity, we assume the usual Hintikka-style axiom schemata for BEL [27] (corresponding to a "weak S5" modal logic). These properties are described in a technical report [15] and will not be reviewed here. Nevertheless, it is worth reiterating the analysis of goals and commitments because their properties figure crucially in speech act analysis.

4.2.1 Goals

At a given point in a course of events, agents choose worlds they would like most to be in — ones in which their *goals* or *choices* are true. (GOAL x p) is meant to be read as p following from the agent's goals/choices, or as p being true in all worlds that are compatible with the agent's goals/choices. Since agents choose entire worlds, they choose the logically and physically

[8] For an exploration of the issues involved in the question of explicit versus implicit belief, see elsewhere [29].

necessary consequences of their goals. Moreover, they choose the expected consequences of their goals — the ones they believe follow from their goals. However, intention will involve a form of commitment that will rule out such expected consequences as being intended.

GOAL has the following properties:

Proposition 1 *GOAL:*
a). $\models \sim(\mathsf{GOAL}\ x\ \mathit{False})$
b). $\models (\mathsf{GOAL}\ x\ p) \wedge (\mathsf{GOAL}\ x\ p \supset q) \supset (\mathsf{GOAL}\ x\ q)$

That is, goals are consistent and closed under consequence. We also have a necessitation property:

Proposition 2 *If* $\models p$ *then* $\models(\mathsf{GOAL}\ x\ \Box p)$

That is, if p a theorem, it is true in all chosen worlds at all times. However, agents do not intend to achieve such "trivial" goals — they are already true.

Unlike BEL, GOAL needs to be characterized in terms of all the other modalities. The semantics of GOAL given in [15] specifies that worlds compatible with an agent's goals must be contained in those that are compatible with his beliefs. This is reflected in the following property:

Proposition 3 $\models (\mathsf{BEL}\ x\ p) \supset (\mathsf{GOAL}\ x\ p)$

If an agent believes p is true now, he cannot now want it to be currently false; agents must accept what they cannot change. Of course, the agent could want it to be false in the future. Conversely, if p is now true in all the agent's chosen worlds, the agent does not believe it to be currently false.

Consider propositions about the future of the form $\Box q$. If the agent believes q is forever true (an example would be a tautology), the above proposition asserts that q is hence forth true in any worlds selected by the agent. Conversely, let p be of the form $\Diamond q$. If the agent chooses worlds in which q will be true sometime in the future (i.e., $(\mathsf{GOAL}\ x \Diamond q)$), the agent cannot have chosen that it be forever false (by Proposition 1a), and hence from the contrapositive of Proposition 3, he does not believe it will be forever false (i.e., $\sim(\mathsf{BEL}\ x \Box \sim q)$) This property re-emerges in our discussion of the preparatory conditions on requests.

Goals are closed under expected consequence. This is easily seen in that, if one believes $p \supset q$, then, by Proposition 3, one has that as a goal as well. Hence, if p follows from one's goals, q does as well.

Finally, we assume that one does not keep achievement goals forever; either one finally achieves them, or one gives up. That is, we assume the following:

Assumption 1 $\models \Diamond \sim(\mathsf{GOAL}\ x\ (\mathsf{LATER}\ p))$

At this point, we have finished summarizing the foundational level, having briefly described agents' beliefs and goals, events, and time. Further discussion can be found in Chapter 2 and in [15]. We now proceed to the characterization of persistent goals and commitment in terms of the preceding concepts.

5 Persistent Goals

In [15], we defined a concept called a P-R-GOAL, for a *persistent, relativized goal* that expressed that way in which an agent is committed to his goals. We then defined intention as a persistent relativized goal to have done a certain action. These two definitions are reiteratd here:

Definition 7 *Relativized Persistent Goals:*

$$\begin{aligned}(\text{P-R-GOAL } x\ p\ q) \stackrel{\text{def}}{=} (&\text{GOAL } x\ (\text{LATER } p)) \wedge (\text{BEL } x \sim p) \wedge \\ &(\text{BEFORE } [(\text{BEL } x\ p) \vee (\text{BEL } x \Box \sim p) \vee (\text{BEL } x \sim q)] \\ &\quad \sim(\text{GOAL } x\ (\text{LATER } p)))\end{aligned}$$

That is, a necessary condition to giving up a P-R-GOAL is that the agent x believes it is satisfied, or believes it is impossible to achieve, or believes $\sim$q. Such propositions q serve as a background that justifies the agent's intentions. In many cases, such propositions constitute the agent's *reasons* for adopting the intention. For example, an agent could adopt the persistent goal to buy an umbrella because of his belief that it will rain. This agent could consider discarding his persistent goal, should he come to believe that the forecast has changed. If q is a proposition of the form (GOAL x $\Diamond$r), then p is a subgoal of $\Diamond$r for the agent. Thus, if x drops r for some reason (e.g., as unnecessary), he can drop p too. If q is a proposition of the form (GOAL y $\Diamond$p), then x is committed to p relative to agent y's wanting p to become true. If x comes to believe that y does not want p to become true, x can abandon his commitment.

Finally, intention was defined as follows:

$$\begin{aligned}\textbf{Definition 8 } (\text{INTEND}_1\ x\ a\ q) \stackrel{def}{=} (&\text{P-R-GOAL } x \\ &[(\text{DONE } x\ (\text{BEL } x\ (\text{HAPPENS } x\ a))?;a)] \\ &q)\end{aligned}$$

Intention is thus a commitment, relative to a background q, to having done an action, believing that one was about to do it.

5.1 Summary

The main advance in the theory of rational action is the development of the concept of a persistent goal, which serves as a foundation for analyzing an agent's intentions and commitments. Although agents are persistent, they are not infinitely so; eventually they give up pursuing their goals.

This ends our discussion of single agents. We now proceed to discuss rational interaction and communication.

6 Communicative Action as Rational Interaction

The formal theory of rational action can now provide a foundation upon which to erect a theory of communicative acts. The process for doing so will go as follows. First, we need

to characterize cooperative interaction sufficiently to deal with a simple request. Then we describe the results of uttering sentences with specific "features" [23], such as utterance mood. Next, we define a general schema for abstracting illocutionary acts and explain how requests can be defined based on certain effects of utterance events. Finally, we show how Searle's conditions on requesting are included in the schematized chain of effects.

6.1 Properties of Cooperative Agents

We describe agents as sincere and helpful. Essentially, these concepts capture constraints (quite simplistic ones) on influencing someone else's beliefs and goals, as well as on adopting the beliefs and goals of someone else as one's own. More refined versions are certainly desirable. Although both concepts are independent of the use of language, ultimately we expect such properties of cooperative agents, embedded in a theory of rational interaction, to provide formal descriptions of the kinds of conversational behavior Grice [22] describes with his "conversational maxims."

First, we shall say that an agent is SINCERE with respect to some other agent y and p, if whenever x has chosen to do something next in order to cause y to believe p, x has chosen to bring it about that y knows p.

Definition 9 (SINCERE x y p) $\stackrel{\text{def}}{=}$

$\forall$ e (GOAL x (HAPPENS x e;(BEL y p)?)) $\supset$ (GOAL x (HAPPENS x e;(KNOW y p)?))

SINCERE is nothing more than an implication; it can be true at some times and false at others. For example, an agent x would be insincere to y about p if x wants y to believe p, and x wants p to be false.[9] Notice that an agent would be insincere if he wants to produce a false belief in another agent, even though he may not believe that he will be successful. That is, as far as we are concerned, insincerity is a matter of the agent's [chosen] desires, not his beliefs.[10] This characterization of insincerity in terms of the agent's wanting to induce false beliefs differentiates our approach from Perrault's [31].

To illustrate the difference between a theory of sincerity based on agents' [chosen] desires and one based on agents' beliefs, imagine an agent who sabotages a nuclear power plant (or even orders a henchman to do so) by making its primary sensor always gives the wrong

[9] Because the definition of sincerity involves making something true, sincerity is "forward-looking" in time (although the event in question can be the empty sequence). The reason for this temporal dimension is that, without it, no performative utterances would be sincere. Briefly, performatives are analyzed as indicative mood utterances about what the speaker has just done. In other words, they are temporally indexical. The analysis of performatives will say that after having uttered such a sentence, the speaker believes he has just done the named illocutionary act. Typically, *prior* to uttering a performative, the speaker has not *just* performed that speech act, and so he would believe his having just done so is false. So, if sincerity involved only what the speaker believed to be true prior to the utterance, no performatives would be sincere. The above definition allows the speaker to sincerely want to do something to get the hearer to believe he has just done the named illocutionary act. For more details, see elsewhere [18].

[10] Of course, these chosen desires must obey the usual constraints. So, the agent cannot believe he will definitely fail to induce a false belief in the agent and yet choose to induce that belief.

reading. Being a good saboteur, the agent decides never again to be in the vicinity of the sensor, and so he never again (if he ever did) knows the sensor's value. That is, he has no beliefs that the sensor has a particular value. However, in rigging up the sensor as he does, the saboteur wants whoever reads it to have false beliefs. One would surely want to say that such a saboteur is insincere, even though he has no beliefs about the facts; such is the nature of sabotage.

In summary, insincerity involves wanting others to come to believe false things, and is a notion independent of language.

Next, consider an agent to be HELPFUL to another agent and an action if he adopts as his own intention the other agent's goal that he eventually do that action (provided that that potential goal does not conflict with his own). Moreover, the agent adopts the intention relative to the other's goals. Should the other agent change his mind, the first agent could nullify his persistent goal.

Definition 10 $(\mathsf{HELPFUL}\ x\ y\ a) \stackrel{\text{def}}{=}$

$$\Box\ (\ [(\mathsf{BEL}\ x\ (\mathsf{GOAL}\ y\ \Diamond(\mathsf{DONE}\ x\ a))) \wedge \sim(\mathsf{GOAL}\ x\ \Box\ \sim(\mathsf{DONE}\ x\ a))] \supset [\mathsf{INTEND}_1\ x\ (\mathsf{GOAL}\ y\ \Diamond(\mathsf{DONE}\ x\ a))])$$

Notice that HELPFUL is defined in terms of $\Box$ (i.e., "henceforth"), meaning once an agent is helpfully-disposed towards another with respect to a specific action (type), he is helpfully disposed from then on. However, this does *not* mean the agent must take on the other agent's goals whenever he believes the other agent wants him to do so. For example, just because an entrepreneur asks a tycoon for $1000 (and gets it) does not mean the tycoon must form the intention to give the entrepreneur more money when asked. In fact, the tycoon could be in a state in which he never wants to give that fellow any more money (because he squandered it) and so never again forms the intention to do so. The second conjunct of the antecedent thus blocks intention formation for these cases.

At this point, we are ready to apply the action and interaction theories for communication.

7 The Effects of Utterance Events

All theories of speech acts and of natural-language communication need to consider the contribution of utterance mood to the effects of the utterances themselves. We assume that the effects of mood will apply to something like propositional content, which we assume for purposes of this paper to be determinable independently of the theory of rational interaction that we are about to develop. Clearly this latter assumption is simplistic; for example, any account of the interpretation of referring phrases or word sense disambiguation, must consider the speaker's intentions [3,11,14,33,26]. However, our strategy is first to develop the theory of rational interaction, then to apply it to sentence-level phenomena before making the transition to problems of utterance interpretation.

It is well known that the form of an utterance does not determine its illocutionary force uniquely. For example, the same imperative utterance could be used to make a request or to issue an order or command. It may not even be used to perform an illocutionary act

at all. Utterance mood is therefore inadequate as an "illocutionary force-indicating device", contrary to its use by S&V [42]. However, given a context, utterance's mood, contributes to the understanding of a speaker's intent. We regard that contribution as a core effect from which many inferences can be drawn. Our concern here is in specifying a logic in to support such inferences, and in describing the core effects.

7.1 Utterance Mood: The Case of Imperatives

Utterance mood conveys a speaker's mental state. In this connection, consider imperatives. The utterance of an imperative to perform some action conveys the speaker's [chosen] desire that the hearer carry out the request, provided the speaker is not thought to have been insincere, Typically, the speaker is also trying to get the hearer to commit himself to the action so that, if all goes well, he does it.

Let us begin to formalize this property of imperatives.

Imperative Property: After speaker x's imperative to addressee y to do action a, if y does not think that x was insincere about his wanting y to do a, i.e., if y does not believe that x wanted y to believe falsely that x wants y to do a, then y believes that x wants y to do a.

To cast the above in our notation, an expression of the following form would be employed:

$$\Box[\,(\mathsf{DONE}\ x\ p?;e) \supset (\mathsf{DONE}\ x\ e;\ (q \supset r)?)]$$

That is, if event e just happened, when p held, then event e was just done, and $q \supset r$ holds. With respect to imperatives, proposition p would restrict e to be the uttering of an imperative sentence in a context in which y was the addressee, x the agent, x and y are attending to one another, etc. Condition q would include the hearer's not thinking that just prior to doing e the speaker was insincere in wanting the hearer to do the act. Finally, r would express the hearer's believing that the speaker wants the hearer to do the act. More formally,

- p would be the uttering of an imperative sentence in the right physical circumstances,
- q would be $\sim$(BEL y [DONE x $\sim$[SINCERE x y (GOAL x $\Diamond$(DONE y a))]?;e]). In words, this means y does not think that x has just done e when he was insincere (i.e., wanted y to come to believe falsely) that x wants y to do action a in the future.
- r would be (BEL y (GOAL x $\Diamond$(DONE y a)))

Now, a simple application of modus ponens shows the following:

Proposition 4 $\models$ (BEL y (GOAL x $\Diamond$(DONE y a))) $\wedge$ (HELPFUL y x a) $\wedge$
$\sim$(GOAL y $\Box$ $\sim$(DONE y a)) $\supset$
(INTEND_1 y a (GOAL x $\Diamond$(DONE y a))

That is, if y thinks x wants y to do a, and y is helpful, and y does not want not to do a, then y will adopt an intention to do a relative to x's desire. Typically, the action a will have a temporal qualification, so the third conjunct of the antecedent would mean that y does not want not to do a by the requisite time.

7.2 Other Effects of Imperatives

Now let us consider what versions of the Imperative Property should hold as we examine different embeddings of what the hearer thinks the speaker believes, as a replacement for ~(BEL y (DONE [~(SINCERE x y (GOAL x ◇p))]?; e) in that property, where p is ◇(DONE y a). First, the conclusion of that property can be reached if we replace the above by ~(BEL y (BEL x (DONE [~(SINCERE x y (GOAL x ◇p))]? ;e), provided that y thinks x is never wrong about his own goals.

Now, if y happens to believe that x did not want y to do the action, and hence that he is insincere, then the conclusion of the Imperative Property does not hold. However, y might still think that x believes his insincerity has not been noticed. Therefore, y may be in the following state:

(BEL y (BEL x [BEL y (DONE [SINCERE x y (GOAL x ◇p)]? ; e)]))

Under these conditions, y might think that x believes y would cooperate and thus attempt to achieve p. An analogous property holds at each level of embedding of (BEL y (BEL x ...)).

We claim that each of these levels is generated by an imperative, provided that there is no corresponding belief in the speaker's insincerity. That is, the hearer jumps to the conclusion that the speaker is sincere as long as there is no belief to the contrary. To summarize, we propose the following: Let e be an event (type) of uttering an imperative sentence, and let q $\stackrel{\text{def}}{=}$ (DONE [~(SINCERE x y (GOAL x ◇p))]?; e), i.e., q is x's having done the utterance event e being insincere to y about his wanting p to become true.

IF AFTER e	THEN
~(BEL y q)	(BEL y (GOAL x ◇p))
~(BEL y (BEL x q))	(BEL y (BEL x (GOAL x ◇p)))
~(BEL y (BEL x (BEL y q)))	(BEL y (BEL x (BEL y (GOAL x ◇p))))
and so on.	

We can express all of these properties at once by using the concept of alternating belief, ABEL. Below we define this concept, develop a notation for characterizing utterance events, and define some domain predicates to use for communication. Then we utilize all of these to formulate a precise statement regarding the effects of imperatives.

7.2.1 Alternating Belief and Mutual Belief

The following defines the auxiliary concept of alternating belief to some level n between two agents x and y that p holds.

Definition 11 (ABEL n x y p) $\stackrel{\text{def}}{=}$ $\underbrace{(\text{BEL x (BEL y (BEL x } \ldots \text{(BEL x p}}_{n}\underbrace{)\ldots)}_{n}$

For example,

If n is	(ABEL n x y p) is
1	(BEL x p)
2	(BEL x (BEL y p))
3	(BEL x (BEL y (BEL x p)))
etc.	

That is, ABEL characterizes the nth alternating belief between x and y that p, built up "from outside in," i.e, starting with x's belief that p. On this basis, one can define unilateral mutual belief — what one agent believes is mutually believed — as follows:

Definition 12 (BMB x y p) $\stackrel{\text{def}}{=}$ $\forall$ n (ABEL n x y p)

In other words, (BMB x y p) is the infinite conjunction[11] (BEL x p) $\wedge$ (BEL x (BEL y p)) $\wedge$ .... Based on the introspective properties we have assumed for beliefs one can show the following is true:

Proposition 5 (BMB x y p) $\supset$ (BMB x y (BEL x p))

Furthermore, from Proposition 5 and from the fact that (BEL x p) $\supset$ (GOAL x p), one easily can show that:

Proposition 6 (BMB x y (BEL x p)) $\supset$ (BMB x y (GOAL y (BEL x p)))

These properties will be useful when we describe the intended effects of imperatives when used as requests.

Before turning to the definition of illocutionary acts, we first add some domain predicates to allow us to specify an illocutionary act's propositional content, then develop notation for describing the effects of utterance events compactly.

7.3 Some Domain Predicates

To have something to communicate, let us introduce a few domain predicates for the logic:

(CLEAN f) — f is clean.
(FLOOR f) — f is a floor.
(DOOR d) — d is a door.
(OPEN d) — d is open.

Next we introduce a few predicates that are true of events:

(FLOORWASHING e) — e is an event of washing a floor.

[11] Barwise [6] has shown that such an infinite conjunction is strictly weaker than a fixed-point definition of mutual belief, such as (BMB x y p) $\stackrel{\text{def}}{=}$ (BEL x p $\wedge$ (BMB y x p))).

(DOOROPENING e) — e is an event of opening a door.

Next, we supply predicates that characterize the semantics of declarative and imperative sentences. Since the development of a full semantic theory lies beyond the scope of this paper, we shall content ourselves with a simplistic version thereof. To say that a *natural-language* sentence is true, we use

(TRUE s).[12]

Next we add a predicate that relates imperative sentences to the properties of events they describe:

(FULFILL-CONDS s e) — Event e fulfills the satisfaction conditions imposed by sentence s.

Clearly, FULFILL-CONDS is just a placeholder for a semantic theory that can characterize the meanings of imperatives. The only requirement we make for analyzing imperatives is that such a semantic theory have the capacity to supply predicates (or properties) that are true of events, especially the utterance event itself (in order to handle performative sentences).

We assume a long list of conditions of the following form:
∀ e (FULFILL-CONDS "Wash the floor" e) ≡ (FLOORWASHING e)
∀ e (FULFILL-CONDS "Open the door" e) ≡ (DOOROPENING e).[13]

7.3.1 Properties Specifically related to Communication

We add to our language the following nonlogical predicates.

(IMPERATIVE s)	s's dominant clause is an imperative.
(DECLARATIVE s)	s's dominant clause is a declarative.
(INTERROGATIVE s)	s's dominant clause is a yes/no interrogative.
(UTTER y s e)	e is a sequence of events in which s is uttered by the agent of e to addressee y.
(ATTEND x y)	x is attending to y.

7.4 Notation for Describing Utterance Events

We shall now define a notation that can be used for declaratives, imperatives, and interrogatives (although in this paper we analyze only imperatives). The purpose of the notation is to factor out the conditions on utterance events necessary for any effects to be realized. Analogous to Searle's [41] "normal input/output conditions," they specify who is speaking (x), who is observing the speech event (y), who is being addressed (z), and what kind of sentence has been spoken (indicated by Φ).

[12] Actually, we should be talking about the truth of *statements*, which resolve the indexicals in the corresponding sentence. Any formulation of a substantive theory for determining which statement is conveyed by a sentence is beyond the scope of this paper. The reader should therefore merely assume that the above makes sense.

[13] Subsequent conditions will bind the agent of the event to be the same as the addressee of the imperative.

Definition 13 $\Phi \Rightarrow \alpha \stackrel{def}{=}$
$\forall$ x, y, z, e, s, n

(ABEL n y x [DONE x ((ATTEND y x) $\wedge$ (UTTER z s e) $\wedge$ (Φ s))?; e]) $\wedge$

$\sim$(ABEL n y x [DONE x $\sim$(SINCERE x z α)?; e]) $\supset$

(ABEL n y x (DONE x e; α?))

That is, $\Phi \Rightarrow \alpha$ is an implication roughly to the effect that if someone believes that utterance event e was just done, where e is the uttering of a sentence s in syntactic mood Φ, and that person does not believe e was done insincerely holding certain "core attitudes" α associated with utterances of that type, then the person believes condition α holds. Because we are dealing with utterance events, the contextual preconditions include what has been termed the "normal input-output conditions" [41], as well as conditions that depend on specific utterance forms. The conditions include: (1) at each level n of alternating beliefs between agents x and y, x was the agent of e, y was attending to x, e is an event of x's uttering sentence s to z, and predicate Φ held of sentence s, and (2) *after* e, it is not the case that at each level of alternating beliefs, the speaker is not thought to have been insincere about α in his performing that event e (i.e., it is thought after the act that the speaker was insincere before the act). If conditions (1) and (2) hold, then at each level of alternating belief α holds.

The utility of the notation is to suppress any mention of condition (1) because it is fixed for all kinds of utterance events. However, α depends on the kind of utterance used, as characterized by Φ.

It should be noted that we do not supply propositions as arguments to primitive utterance events, as is done in [1,13,32,31] because doing so requires one to characterize semantically how events can operate on functions from possible worlds to truth values (or whatever is one's semantic analysis of propositions). We do not know how to do this. By talking explicitly about sentences, we have a hope of integrating a semantic theory with a theory of utterance effects.

Given this notation, the following domain axiom is used to characterize an imperative utterance.

7.5 Characterizing Imperatives with the New Notation

First, let p $\stackrel{def}{=}$ $\exists$e' (DONE z e') $\wedge$ (FULFILL-CONDS s e')

That is, p says that some event e' has just been done by y and satisfies the fulfillment conditions of sentence s. To describe the context-dependent effects of imperatives, we employ the following:

Domain Axiom 1 *Imperatives:*

$\models$ IMPERATIVE $\Rightarrow$ (GOAL x $\Diamond$p)

That is, if an imperative sentence s to make p true (i. e., to make it the case that there be some event e' that fulfills the conditions given by sentence s) is uttered in circumstances in which neither speaker nor hearer suspects the speaker is insincere in his wanting the hearer to make p true,[14] the hearer thinks (and thinks the speaker thinks, etc.) that the speaker wants p to become true — i.e., that the speaker wants the hearer to carry out an action fulfilling the conditions specified in the utterance.

A noteworthy aspect of the above property is that the level-counting variable n is quantified across both sides of the implication.[15] It therefore picks out each of the above levels of alternating belief. At level n = 1 of this axiom, if after the utterance of an imperative, y thinks that x was insincere about his goal that z do something, then y need not believe, that x wants z to do it. However, at level n = 3, y could believe that x thinks that y believes x was sincere, so y could think that x believes that after uttering the imperative, y will believe that x wants z to act.[16]

The same argument can be made for any other levels of embedding of x believes that y believes, etc. Hence, any level of alternating belief that the speaker is insincere about wanting y to do the specified action will nullify the conclusion *at that level* of alternating belief, but not at any others. The ability of ABEL to characterize arbitrary depths of alternating beliefs (regarding insincerity and the speaker's goals,) in this case allows *one* axiom schema to capture ironic and insincere imperatives as well as the usual case.

Since $\forall w\,(P(w) \supset Q(w))$ implies $\forall w P(w) \supset \forall w Q(w)$, as we quantify over the positive integers indicating levels of alternative belief, we can derive the conclusion that y thinks it is mutually believed (in our notation, BMB'ed) that the speaker x wants y to achieve p. Illocutionary acts will be defined to require that the speaker intend to produce such beliefs about mutual beliefs, but it is important to notice that utterance events will not lead to mutual beliefs if performed in circumstances in which the speaker is suspected (at some level) of insincerity. Those cases in which there is a suspicion of insincerity will constitute a "defect" in the performance of an illocutionary act. Although an illocutionary act defective in this way can still succeed, one might not want to classify the utterances in question as a full-blooded illocutionary act. The hearer's deciding to cooperate and attempting to execute the action depends only on what the hearer actually thinks the speaker wants and intends.

[14] We have taken this position as a result of Ray Perrault's criticism in this volume of the "attitude-independent" analyses of locutionary acts presented by us at the 1985 ACL conference [16]. Without his criticism and suggestions for an approach based on default logic, our analysis would be more complex than it is. Although we differ on substantive technical issues, such as the use of default logic, the two approaches now have much in common.

[15] Notice that x is a free variable here, and is intended to be captured by the definition of $\Rightarrow$.

[16] Note, by the way, that x could be the same as y, and that y could be the same as z. Since we assume that agents are always attending to themselves, the axiom characterizes effects of uttering an imperative on speaker, listeners, and addressees simultaneously.

Last, the propositional content of an imperative is captured by using FULFILL-CONDS, which maps the imperative sentence to the conditions on some future event that the speaker wants to hold.

With this understanding of the effects of imperatives, we can proceed to the first part of our method for characterizing illocutionary acts: infer from utterance form to illocutionary point. The next step is to derive some of the effects of an imperative, and then characterize an attempt to achieve something. We then define an illocutionary act as an attempt to achieve certain effects.

8 The Effects of Imperatives

Given no insincerity, the uttering of an imperative by speaker x to addressee y to do action a results in y's thinking it is mutually believed that x wants y to do a. As a consequence of Proposition 4, if y does not mind doing the action, and is helpfully disposed towards x and a, then we can conclude that y intends to do the act relative to x's desire. From Theorem 1 in [15], we can identify the conditions guaranteeing that an intention to act will in fact be achieved. Let us call those conditions world-right, then

$$(\mathsf{INTEND}_1\ y\ a\ q) \wedge \mathsf{world\text{-}right} \supset \Diamond(\mathsf{DONE}\ y\ a)$$

Thus, given an expression of the speaker's desire that the hearer act, and the hearer's helpfulness, lack of objections to doing the act, and the appropriateness of the world, we can conclude that the act happens. This is the usual and desired reasons for planning to issue an imperative utterance (as a directive).

Now, not only do speakers want this chain of effects to take hold, when they make public their desires about the initial parts of the chain, under certain circumstances, they make public their desires for the rest. To see this, we point out that the following holds:

Proposition 7 $(\mathsf{BMB}\ y\ x\ (\mathsf{GOAL}\ x\ \Diamond(\mathsf{DONE}\ y\ a))) \wedge (\mathsf{BMB}\ y\ x\ (\mathsf{GOAL}\ x\ (\mathsf{HELPFUL}\ y\ x\ a))) \wedge$
$(\mathsf{BMB}\ y\ x\ (\mathsf{GOAL}\ x \sim(\mathsf{GOAL}\ y\ \Box \sim(\mathsf{DONE}\ y\ a)))) \supset$
$(\mathsf{BMB}\ y\ x\ [\mathsf{GOAL}\ x\ (\mathsf{INTEND}_1\ y\ a\ [\mathsf{GOAL}\ x\ \Diamond(\mathsf{DONE}\ y\ a)]\ )])$

So, if an agent makes the hearer think it is mutually believed he wants the hearer to do something, and the hearer thinks it is mutually believed that he wants the hearer to be helpful,[17] then the hearer has made the hearer think it is mutually believed that the speaker wants him to intend to do the action relative to the speaker's [chosen] desire. The proof of the above follows from Propositions 5 and 6, along with Proposition 4 and consequential closure of BMB and GOAL.

Similarly, one can extend the line of inference to show that if the hearer thinks it is mutually believed that the speaker wants the "world to be right," then the hearer thinks it is mutually believed that the speaker wants the act to be done eventually. That is, we have

[17] Just how speakers make that true is an interesting issue, related to politeness, mitigation, uses of "please," tone of voice, etc., but not an issue we can address here.

embedded the earlier line of inference within (BMB y x (GOAL x ...)). Given all these effects, what shall we say a speaker characteristically attempts to achieve when issuing a request? To answer this question, we first need to discuss attempts.

9 Attempts

A crucially important property of illocutionary acts, not shared by some other actions (e.g., knocking over a glass of water, or simple utterance events) is that they cannot be performed accidentally or unknowingly. To illustrate this, consider the following example.[18] A blindfolded person reaches into a bowl of flashcards, pulls out three cards, and knowingly turns them towards another person. The cards say "Open the door." One would not be inclined to say that a request to open the door took place, in part because the agent was not committed to conveying that specific content. To exclude such cases from being labeled as communicative action, we first want to examine only actions in which the agent is committed to performing the action, and perhaps to achieving certain states-of-affairs. Actions so performed are not accidental.

As we have seen earlier, there are at least two kinds of states an agent can be in with respect to a chosen desire, say p. First, and most strongly, the agent can be committed to achieving p; if the agent fails, we would expect him to try again, all else being equal. This is expressed with P-R-GOAL and INTEND. A second, weaker, mental state is to want achieve p, but not be committed to achieving it. We express this mental state with the conjunction of the agent's believing p is currently false, and his wanting it to become true (next). But, in this state, if the agent fails to achieve p, no prediction about a second attempt would be made.

Now, attempts can involve either of both of these types of goal states. To capture this, we specify a simple form of an *attempt* as follows:

Definition 14 {ATTEMPT x e p q r} $\stackrel{def}{=}$

$$\begin{array}{ll} (& [\mathsf{INTEND}_1\ x\ e;p?\ r] \wedge \\ & [(\mathsf{BEL}\ x \sim q) \wedge (\mathsf{GOAL}\ x\ (\mathsf{HAPPENS}\ x\ e;q?))]\ )? \quad ;e \end{array}$$

That is, e is an attempt if it is done in the right circumstances, namely when the agent (1) intends e should achieve p relative to background r, and (2) has chosen that e should also achieve q. Of course, either of these formulas could be trivialized by substituting the proposition *True* for the appropriate schematic variable. In summary, the definition above factors out commitment from achievement in an agent's attempt.

To illustrate the above definition, consider attempting to sink a game-winning, last shot in a basketball game. The event e in question is a certain movement of the agent's body. The agent is committed to shooting (e), and hence to the ball's being launched unimpeded toward

[18] We are indebted to Ray Perrault for this example.

the basket (p); the agent wants to achieve the ball's being in the basket (q) Thus, the agent's commitments to doing the action and to p characterize his doing his part, and the rest is up to physical and causal laws.

10 Definition of Request

To characterize a request or, for that matter, any illocutionary action, we must decide on the appropriate formulas to substitute for p, q, and r in the definition of an attempt. Two observations guide our formulation. First, Allen and Perrault's work [2,32] shows that the key to a formal analysis of indirect speech acts is to understand that there may be many routes leading to the conclusion that any particular effect (here the effect of an imperative) holds. Thus, an action for uttering an imperative will not be included as part of requesting, but instead some unspecified sequence of events will be used.

Second, we shall use the aforementioned effects of an imperative to specify what the speaker was attempting to achieve. Ultimately the question of what effects should be encapsulated in a complex action expression can be settled only on empirical and philosophical grounds. For example, each choice of intended effect entails certain gating conditions that then become part of the speech act definition. Moreover, by stipulating that a certain effect is *intended*, rather than merely chosen, one states that the agent is committed to achieving it; if he does not, we predict that he will try again. Furthermore, the agent believes that the effects he attempting to achieve are not already true. Thus, the felicity conditions to which the theorist is committed are determined by the theorist's choice of the commitments attributed to speakers in performing instances of various kinds of illocutionary acts. We argue in favor of one choice below, but it should be emphasized that what has been developed here is actually a framework for formulating many theories.

Let us now define a request by incorporating some of the effects of imperatives. The core effect from which we start is:

(BMB y x (GOAL x ◇(DONE y a)))

From this, we can (as mentioned above) derive, under the right conditions, the consequences of Propositions 4 and 8, i.e.,

(INTEND_1 y a (GOAL x ◇(DONE y a))), and

(BMB y x (GOAL x [INTEND_1 y a (GOAL x ◇(DONE y a))]))

We now incorporate these effects into a complex action expression called REQUEST.

Definition 15 {REQUEST x y e a r} $\stackrel{\text{def}}{=}$
{ATTEMPT x e [(BMB y x (GOAL x ◇(DONE y a))) ∧

$$
\begin{aligned}
&(\mathsf{BMB}\ y\ x \\
&\quad (\mathsf{GOAL}\ x \\
&\qquad (\mathsf{INTEND}_1\ y\ x\ [(\mathsf{GOAL}\ x\ \Diamond(\mathsf{DONE}\ y\ a)) \wedge (\mathsf{HELPFUL}\ y\ x\ a)]))) \\
&(\mathsf{INTEND}_1\ y\ x\ [(\mathsf{GOAL}\ x\ \Diamond(\mathsf{DONE}\ y\ a)) \wedge (\mathsf{HELPFUL}\ y\ x\ a)])
\end{aligned}
$$

That is, event e by speaker x to hearer y to do action a is a request (relative to background r) iff in performing e x's goal is committed to making y think it is mutually believed that (1) x wants y to do a, and (2) x wants y to intend to do a because x wants y to do it, and because y is helpful.[19] Lastly, in making a request, the speaker is committing himself to make public his desire that the hearer carry out some act a, and he also wants thereby to achieve the hearer's forming a commitment to act in accordance with the speaker's expressed desire, because the hearer is helpful.

A minimal commitment for the performance of an illocutionary act is the speaker's commitment to making it mutually believed that the speaker is in a certain mental state with respect to the content of the utterance. For example, if the speaker asks a hearer to open the door and the speaker learns subsequently that his voice was garbled, we can predict that the speaker will try again to make his chosen desires "public" to the hearer.[20] Thus, the assumption that speakers are performing illocutionary acts [5], that is, that speakers are trying to communicate, entails the assumption that the speaker is committed to conveying that he is in a certain mental state. As remarked upon by Winograd and Flores [44], conversation implies commitment.

The second commitment is necessary to rule out the following from counting as a request: A particularly nasty individual says to Oedipus "I want you to marry your mother." In saying this, the speaker intends to convey (1). Of course, both agents know the hearer will not be helpful, and will not form the intention, because no one would intentionally marry his mother. Hence, if the speaker is judged to be sincere, he must want the hearer to marry someone but not know she is his mother, and this is mutually known. Hence, this is clearly, not a request that Oedipus marry his mother. In summary, to block mere expressions of desire from being labelled as requests, we make the definition of requesting incorporate a speaker's being committed to making public that he wants the hearer to form an intention.

In addition to that commitment, the speaker intends to make it mutually believed that he wants the hearer to form the intention to act, relative to the speaker's desires and to the hearer's helpfulness. Furthermore, the speaker intends it to become public that he wants the hearer to form an intention *relative to the speaker's desires.* If the speaker later says "never mind," the hearer can drop the commitment. This helps to explain why, after acceding to a request, say with "OK," the addressee has performed a commissive speech act. In addition,

[19] We are reading "causes" and "reasons" into the background condition, but more precisely, instead of saying p is a reason or a cause for a goal or intention, we should say the goal or intention is required to persist as long as p is true.

[20] Of course, the formalism also allows the speaker to give up his attempt if he thinks nothing he can do will make the transmission of his desires any more successful.

the hearer is supposed to form that intention relative to his being helpfully disposed towards the speaker (and the action). That is, helpfulness should be one of the reasons the hearer forms the intention to act (for a request); for a command, one of the reasons should be the speaker's having authority over the hearer. However, we have had to stipulate, rather than derive, that a request requires that the speaker wants it to be mutually believed that he wants the hearer to form the intention to do the action *because* the hearer is helpful. That is, we have stipulated that the speaker wants HELPFUL to be part of the "background" clause for the commitment underlying the hearer's intention. Importantly, given the definition, once the hearer is helpfully disposed towards the speaker and a given action, he is henceforth always so helpfully disposed. Hence, once adopting an intention because one is helpful (and not minding doing it), the addition of HELPFUL to the release clause adds nothing with respect to the hearer's ability to drop the commitment. However, what we are stipulating here is that the speaker *wants* the hearer to form the intention in this way, and the speaker need not have any beliefs about the hearer's helpfulness.

Now, in addition to being committed to making public his desires, the third conjunct of the REQUEST definition states that the speaker wants to achieve the hearer's actually forming the appropriately relativized intention. A request seems incomplete if the speaker intends to convey his desires but does not want the hearer in fact to form the intention to act. (Notice that an informative speech act about the speaker's desires does not seem incomplete). We do not insist that the speaker be committed to the hearer's forming an intention because that would require the speaker to be more insistent about getting compliance than we think is necessary; the speaker should simply be able to take "no" for an answer. Moreover, by not requiring the speaker to be committed to the hearer's forming such a relativized intention, we do not force the speaker to request again when the hearer, believing the speaker has authority over him, adopts an intention for the wrong reason. In making a request, the speaker wants the hearer to be helpful, but does not necessarily insist upon it.

It is important to note that the speech act of requesting requires the hearer to regard it as mutually believed, rather than merely believed to a finite level of embedding, that the speaker wants the hearer to do the action. Thus, felicitous requesting will require that the hearer presume the speaker's sincerity at all levels of alternating belief (though not necessarily the actual fact of sincerity.)

Finally, the above definition works for indirect as well as direct requests. It does not specify *how* the hearer arrives at the mutual belief that the speaker wants the hearer to perform an action. With respect to a direct request, this mutual belief may be an immediate consequence of the speaker's uttering an imperative. But it may also be the result of an inference from some other event(s). For example, if a speaker utters "Get the hammer" to a hearer when the speaker is standing on a ladder and is obviously holding a nail in position, the hearer may infer it as mutually believed that not only should the hearer get a hammer (satisfying the direct request), but he should also hand it to the speaker. This latter mutual belief is sufficient to initiate the [intended] inference path that satisfies the indirect-request interpretation.

10.1 Linguistic Requests

The above definition made no mention of a sentence. Hence, we can have requests made exclusively by gestures, provided that *some* non-linguistic means is available for revealing the speaker's chosen desires. To specify a request by means of language, we define the following:

Definition 16 $\{\text{LING-REQ } x\ y\ e\ s\ e'\ r\} \stackrel{\text{def}}{=} ((\text{UTTER } y\ s\ e) \wedge (\text{FULFILL-CONDS } s\ e'))?;$
$\{\text{REQUEST } x\ y\ e\ e'\ r\}$

In other words, a linguistic request is characterized as a complex action expression that requires e to be an utterring by agent x of sentence s to addressee y. The definition specializes the proposition that y is to make true (in the REQUEST) to be her effecting an event e' that fulfills the conditions imposed by the sentence s.

10.2 The Point

By turning requests into complex action expressions, we can now say when a request has been made, and so we can reason about what would be true if a request were done. However, we have not had to add anything to the formal language. The notion of requesting is entirely meta-theoretic, having been described here with definitions. One could view these definitions as expanded "in line" into their components, but this is not necessary. Furthermore, we do not need to say that communication requires illocutionary-act recognition, since hearers can infer the needed effects and respond appropriately without any such explicit recognition. Simply put, a request is not a new type of primitive event, rather it is an event of some other type that happens to occur in the right circumstances.

How much of a chain of effects should be incorporated into the definition of an illocutionary action? Just as mathematicians have the leeway to decide which results are useful enough to be named as lemmas or theorems, so too does the language user, linguist, computer system, or speech act theoretician have a great deal of leeway in deciding which complex action expressions to form and name. Grounds for making such decisions range from the existence of illocutionary verbs in a given language to considerations of efficiency. However, complex action expressions are flexible — they allow different languages and agents to carve up the same chains of inference in different ways. For example, we have shown the implications of committing an agent to producing an effect; if the agent does not achieve it he will try again. Just how much of any given chain should be defined as constituting the speaker's commitments is a matter of subsequent argument. It is not an essential feature of our analysis.

The complex action expression named REQUEST could thus have been named anything at all. We are not making any claims as to the existence of a general mapping between such action expressions and English illocutionary verbs. There could be long chains of inference incorporated into action expressions for which a particular natural language contains no illocutionary verb. For example, an action expression labeled "want-request" might capture the inference from someone's saying "I want p" to the hearer's bringing p about. This freedom allows agents to create complex action expressions for "conventionalized" indirect uses of sentences, even though there is no verb in the natural language to describe those uses.

10.3 Illocutionary Act Recognition

So what does recognition of an illocutionary act amount to? There are two steps. First, it involves recognizing that the speaker was attempting (and hence intending) to get the hearer to draw certain inferences. For example, after receiving an imperative, the intended addressee must first come to believe mutually with the speaker that the latter wants him to do something. Next, our analysis proposes that the recipient's helpfulness causes him to take on a commitment to carry out the action conforming to that speaker's desire. We do *not* say that the hearer has to adopt that commitment because he was intended to do so (which he was, if the speaker was in fact requesting).

He might in fact form the intention to act because the speaker has authority over him. There is no requirement being proposed that utterances are defective (and the conversational state should be repaired) if they are illocutionarily indeterminate to the hearer. Given such indeterminacy, one might want to say that the speaker incompletely or defectively performed an illocutionary act of a given type. But nothing of importance follows from that fact in our theory of rational interaction. Of course, various social and institutional consequences could follow from fact that the speaker did not completely perform a given illocutionary act, but that involves additional stipulations about the institutional and social realms of interaction that we are not addressing.

Now, *if* the hearer in fact adopts the commitment because he comes to believe that the speaker intended him to adopt it by virtue of his helpfulness, the hearer is then embarked on the path of recognizing what illocutionary act was performed. The hearer would need to achieve complete recognition of the intended line of inference so as to have enough information to identify what kind of illocutionary act was being performed. Moreover, the hearer would have to recognize which effects the speaker was committed to, which he was trying to achieve, which he has chosen, and which are merely fortuitous. But our point is that the hearer need not perform this entire recognition procedure — he may just be helpful and do the action. In summary, the first step of illocutionary-act recognition is to determine which effects the speaker intended. This may or may not be done by a hearer.

The second step is to then examine whatever line of intended inference has been uncovered to determine within what illocutionary act(s) it is encapsulated. In other words, what we have here is a classificatory process. Whereas, for example, the first step may well be useful, for ascertaining *how* one is supposed to adopt one's attitudes, say, by virtue of helpfulness or because of the speaker's authority, the second classificatory step contributes little in principle. From a heuristic standpoint, if one assumes that the intended effects are *all* encapsulated in an illocutionary-act definition, then, if some of those effects can be inferred, then, one may hypothesize an illocutionary-act classification and thereby predict what other effects the speaker may have intended (much as one can parse bottom-up with top-down prediction). Still, there is no reason to believe that speakers intend to achieve all of the effects encapsulated in an illocutionary verb *and only those effects.* The heuristic value of illocutionary act recognition thus remains to be seen. Our main point here is that actual identification adds nothing in principle.

Finally, because it may take a speaker more than one utterance to perform a complete

illocutionary act, and because utterances are often completed by other speakers [12], hearers would have to recognize which illocutionary act was performed after sequences of utterance events. That is, hearers would have to look back arbitrarily far (subject to constraints such as those described by Grosz and Sidner [25]) to see which illocutionary acts the current utterance was completing. Illocutionary act recognition thus seems to us unnecessary, unlikely, and uninformative.

10.4 Summary

We have defined requests in the following way. First, a logic of intention was developed. Next, imperatives were analyzed, since they are the prototypical way in which directive speech acts are performed. The "core" effects of imperatives, the speaker's desire that the hearer should act, are revealed to the hearer provided the hearer does not believe that the speaker was insincere in making his utterance. Then, we derived other important effects, namely that the hearer thinks it is mutually believed that the speaker wants him to intend to act because he is helpful and because the speaker wants him to act. Another effect is that if the hearer is in fact helpful, and does not mind acting, he forms the intention to act relative to the speaker's desire. Finally, we define a request as an attempt to achieve these conditions, which entails certain commitments. In making requesting into a complex action, we abstract away from any primitive utterance event (or sequence of them); instead, the speaker is viewed as having performed a request if he executes any sequence of actions that produces the needed effects.

To show that these principles can provide the basis for an adequate analysis of illocutionary acts, we show how Searle and Vanderveken's [42] conditions on requesting can be derived from an independently motivated analysis of intentional action, as S&V recommend.

11 Deriving Searle and Vanderveken's Conditions on Requesting

Searle and Vanderveken's conditions on speech acts are divided into the *normal input/output*, *propositional-content*, *preparatory*, *sincerity*, and *illocutionary point* conditions. We explore each of these in turn for requests. Where not superseded by [42], we also refer to Searle's *Speech Acts* [41].

11.1 The Normal Input/Output Conditions

The normal input/output conditions include the (1) speaker's and hearer's ability to speak and understand the relevant language, and (2) conditions on the utterance itself, such as audibility. Also included herein would be, presumably, other conditions on the modality of communication, such as distinctive features of speaking on the telephone, computer-mediated conferencing, or text. Finally, normal I/O conditions would exclude "parasitic forms of communication," such as telling a joke or acting in a play.

These I/O conditions are supposed to apply to illocutionary acts *per se*. In our scheme, however, these conditions apply to utterance events themselves, not to illocutionary acts. A request, for example, can be performed even if the participants do not speak a shared language or nothing at all is uttered (though *something* had better be observed, say, a printed word or a gesture). Whereas many of the conditions we propose would be preconditions on any utterance event's conveyance of the speaker's mental state, other conditions depend on the kind of utterance used. Thus, the conditions we state are a function — at least in part — of the utterance event's signaling what it does about the speaker's mental state, as well of the utterance's form and shared beliefs with regard to its meaning.

The illocutionary-act definitions, however, are independent of these conditions. For the speaker to attempt to achieve various effects with an utterance, he must believe that these conditions hold. But, the definition of illocutionary acts depends solely on what the speaker is trying to achieve, not on what he thinks must be true for a specific utterance to achieve it. An utterance event can achieve the effects that are necessary for performing an illocutionary act only if that event occurs under the right conditions; these depend, for example, on the modality of communication. For example, no one would claim that a request uttered over a telephone is a different kind of illocutionary act from one uttered face-to-face.

Finally, "parasitic" forms of communication (e.g., jokes, irony, etc.) are handled in the same way as we treat insincere utterances. Whenever the hearer believes (or believes that the speaker thinks he believes) the speaker is insincere, the effects normally produced by utterances of that kind are correspondingly weakened. Ultimately, when everyone knows the speaker is insincere the usual effects are not produced. We do not say specifically what *is* produced, but that is another matter.

In sum, the normal input and output conditions should apply to the utterance event itself and not be incorporated directly into the definition of illocutionary acts because instances of the same illocutionary act type can be performed with many kinds of utterances.

11.2 The Propositional-Content Condition

Searle's condition states that a speaker requests a future act to be done by the hearer. This condition appears in our definition as the speaker's being committed to making public his chosen desire that the hearer do some future action.

11.3 The Preparatory Conditions

Searle proposes two preparatory conditions, each of which is partially satisfied by our account. First, the hearer should be able to do the requested act. Second, the speaker believes that the hearer can do so. These conditions seem to us to be too strong. We believe that one can make a perfectly good request independently of whether or not the hearer can *in fact* do the requested act. Furthermore, the hearer may not believe he can do it but may believe that the speaker thinks he can. In fact, we think that speakers can make felicitous requests even though the speaker might not be sure that the hearer can in fact do the act. All that appears to be required is that the speaker not believe that the hearer *cannot* do the act. For example,

the point of a request could be, among other things, to confirm that the hearer can perform the act.

Our analysis supports this weaker claim about speakers' beliefs in the following way: First, let us recall that a request is defined to be an event that makes the hearer think it mutually believed that the speaker's goal (chosen desire) is that the hearer eventually carry out a specified action. The semantics of GOAL are such that, if one's goal is $\Diamond p$, one does not believe $\Diamond p$ to be false, i.e., one does not believe that $\Box \sim p$. Hence, the hearer would think it mutually believed that the speaker does not believe the hearer will never act as requested. Of course, since a request is an attempt to make the speaker's goals public, if the speaker is sincere, he actually has the goal that the hearer so act, and thus, he is required not to believe the hearer cannot act. Moreover, since the speaker is also attempting to get the hearer to form an intention to do the requested action, the speaker thinks that the hearer does not believe he will never do the requested action. Thus, the semantics of GOAL plays the key role in satisfying a more accurate version of Searle's first preparatory condition.

The second preparatory condition is that it should not be obvious to either speaker or hearer that the latter was going to do the act "*in the normal course of events and of his own accord.*"[21] In our framework, this amounts to the hearer's already having a persistent goal to accomplish the act, but one that need not be relative to the speaker's desire. To encode its not being obvious that the hearer has such a persistent goal, we could say that the speaker does not believe (or, perhaps, mutually believe) that the hearer already intends to act. But, again, we believe this statement of the non-obviousness condition to be too strong. If the speaker believed the hearer intended to do the act at the time of making the utterance, he could still be attempting to get the hearer to form a persistent goal to act *relative to the speaker's desire.* That is, one purpose of a request is to get the hearer to do something *for* the speaker.

However, if the hearer is *already* committed to an action *for* the speaker, then a second imperative to do that action will constitute not a felicitous request, but perhaps a form of badgering. This is so because requests are, in part, attempts to commit the hearer to the speaker; our attempt definition involves the agent's wanting to *achieve* certain effects, and that achievement requires that the speaker believe those effects to be false. Consequently, the second this form of badgering not be a full-fledged request by our definition.

11.4 The Sincerity Condition

The sincerity condition for a request is that the speaker should want the act done. Sincerity arises immediately from the definition we provide for requesting since the definition involves the speaker's attempting, hence intending, to do some action to get the hearer to think it mutually believed the speaker has a certain goal (chosen desire), namely, that the hearer act. Of course, intentions in our scheme are built out of goals, so we immediately deduce that the speaker's goal is for the hearer to believe something about the speaker's mental state. At this

[21] Searle, [41], p 66. Emphasis is ours. This condition is not present in The Foundations of Illocutionary Logic [42].

point, the antecedent of the implication constituting SINCERE will apply. So we deduce that a sincere request is one in which the speaker indeed wants the hearer to act as requested. For us, a request can be made even if it is ultimately insincere, but not if it is recognized as insincere. This is a theoretical position we have taken, but not one that is forced by the formalism. To define illocutionary acts to require total sincerity, we need only define the act to use mutual knowledge instead of BMB.

11.5 The Illocutionary Point

S&V state that the illocutionary point of a request is that the speech act should be *an attempt* to get the hearer to do a certain act. Following them, this is precisely how we have defined REQUEST.[22] However, most of the other non-directive speech acts described in [41] are not characterized therein as attempts to achieve some illocutionary point. In our opinion, they should be. By defining illocutionary acts as attempts, one can see why only illocutionary verbs can be used as performatives; for these, only the right intentions and beliefs are necessary. This will be shown in a subsequent paper.

We have demonstrated the adequacy of our approach by showing how Searle's conditions on requesting could be derived from principles arrived at independently. We describe briefly how other illocutionary acts can be handled.

12 Other Illocutionary Acts

We have concentrated here on the prototypical directive illocutionary act. The assertive class is analyzed similarly: after a declarative sentence has been uttered, if there was no suspicion of insincerity (at any level of alternating belief), then the hearer thinks it mutually believed that the speaker believes the propositional content. The illocutionary act of assertion is defined as an attempt to achieve a mutual belief (BMB) that the speaker believes the content. The illocutionary act of informing is defined as an attempt to get the hearer to believe the content as a consequence of arriving at this mutual belief about the speaker's belief. Of course, a theory of evidence is needed to describe the conditions under which believing that the speaker believes something should cause the hearer to adopt a similar belief.

What do we have to say about the others, namely Searle's commissives, expressives, and declaratives? First, we have little to say about expressives; one needs to characterize the mental states (e.g., sorrow, regret) they embody. Then one needs to correlate utterance features with the fact that the speaker is in one of these mental states. For the time being, we

[22] Searle and Vanderveken [42] claim illocutionary point determines the rest of the dimensions:

> All general propositional content, general preparatory, and general sincerity conditions are determined by its illocutionary point. The sense in which they are determined is simply that one cannot achieve that illocutionary point without presupposing these preparatory conditions, without expressing these sincerity conditions, and without expressing a proposition satisfying those propositional content conditions. [p. 50]

We agree, and our work can thus be regarded as a formal theory in support of that claim.

shall content ourselves with analyzing expressive speech acts as assertives that the speaker is in the requisite state. If sincere, then he is.

Expressives are frequently conveyed through performative (e.g., "I apologize..."). We can show how performatives can be analyzed without recourse to a separate category of speech act, i.e., the declaration [42]. In this kind of speech act, the speaker makes something true by saying so. We can show how the assertive kind of speech act (coupled with institutionally-based facts) can solve problems of performatives — problems for which S&V propose the declaration speech act type.

Last of all, let us now consider commissives. According to Searle and Vanderveken, uttering a commissive establishes a "commitment." We claim that a necessary condition on the acceptance of an interpersonal commitment is to make it mutually believed with another agent that one has adopted a persistent goal to achieve something relative to that other agent's desires. This relativization of an internal commitment to another's desires shows why a speaker cannot felicitously promise to do something for a hearer that the speaker knows the hearer does not want. Moreover, it shows why an interpersonal commitment has been made in responding positively to a request; the speaker's intention in requesting involves the hearer's taking on just such a relativized commitment.

According to S&V, certain commissives such as promises "strengthen" this commitment so that it becomes an obligation. As we do not analyze obligations here, however, our theory is incomplete — but incomplete in the same way as not having an analysis of regret and sorrow. Still, unlike these mental states, there is clearly much in common between a notion of obligation and our analysis of relativized commitment. Whereas the former is institutional and social in nature, the latter is cognitive. However, we contend that it is the ability to adopt such an internal commitment that makes having an obligation possible.

13 Concluding Remarks

This paper has demonstrated that not all illocutionary acts need be primitive but rather can be treated as complex actions. Many properties of these actions can be derived from more basic principles of rational action and from 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 of communication are independently motivated. In particular, the theory of rational action is developed independently of any notions of communication.
- The characterization of the result of uttering sentences with certain syntactic moods is justified both 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).
- Complex action expressions 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 such an expression for purposes of efficiency, but it need not correspond to that of any other agent.

- The theory provides solutions to problems of performatives.[23]
- The rules for combining illocutionary acts (characterizing, for example, how multiple assertions could constitute the performance of a request) now have been 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 the utterances in question and showing that the combined effects constitute the illocutionary act.
- Multi-act utterances are also a natural outgrowth of this approach. Sentences can be uttered in circumstances that satisfy the conditions of several illocutionary acts.
- The theory is naturally extensible to indirection (to be argued for in another paper), to other illocutionary acts, such as questions, commands, informs, and assertions, and to the act of referring [14].

In summary, we have presented a theory of speech acts and communication grounded in a theory of rational interaction. In so doing, we have sought to demonstrate that there is no need to propose a separate logic for illocutionary acts; the logics of attitudes and action should be entirely satisfactory.

14 Acknowledgments

We would like to express our appreciation to Ray Perrault, who provided valuable criticism and suggestions. Also quite helpful were discussions with Doug Appelt, Michael Bratman, Herb Clark, Jim des Rivières, Lyn Friedman, Barbara Grosz, Jerry Hobbs, David Israel, Kurt Konolige, Joe Nunes, Calvin Ostrum, John Perry, Martha Pollack, Syun Tutiya, and Dietmar Zaefferer. Many thanks to them all.

References

[1] J. F. Allen. *A Plan-Based Approach to Speech Act Recognition.* Technical Report 131, Department of Computer Science, University of Toronto, Toronto, Canada, January 1979.

[2] J. F. Allen and C. R. Perrault. Analyzing intention in dialogues. *Artificial Intelligence*, 15(3):143–178, 1980.

[3] D. Appelt. *Planning English Sentences.* Cambridge University Press, Cambridge, U. K., 1985.

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

[23] Our discussion of performatives in this framework, is now being prepared for publication [18].

[5] K. Bach and R. Harnish. *Linguistic Communication and Speech Acts.* M. I. T. Press, Cambridge, Massachusetts, 1979.

[6] J. Barwise. Three views of common knowledge. In M. Vardi, editor, *Proceedings of the Second Conference on Reasoning about Knowledge*, Morgan Kaufman Publishers, Inc., Los Altos, California, March 1988.

[7] R. Brachman, R. Bobrow, P. Cohen, J. Klovstad, B. L. Webber, and W. A. Woods. *Research in Natural Language Understanding.* Technical Report 4274, Bolt Beranek and Newman Inc., Cambridge, Massachusetts, August 1979.

[8] M. Bratman. *Intentions, Plans, and Practical Reason.* Harvard University Press, 1987.

[9] M. Bratman. Two faces of intention. *The Philosophical Review*, XCIII(3):375–405, 1984.

[10] M. Bratman. *What is Intention?* Volume SDF Benchmark Series, MIT Press, Cambridge, Massachusetts, 1988.

[11] H. H. Clark and C. Marshall. Definite reference and mutual knowledge. In *Elements of Discourse Understanding*, Academic Press, New York, 1981.

[12] H. H. Clark and D. Wilkes-Gibbs. Referring as a collaborative process. *Cognition*, 22:1–39, 1986.

[13] P. R. Cohen. *Chapter 5: The Pragmatics/Discourse Component, Research in Natural Language Understanding.* Technical Report 4274, Bolt Beranek and Newman, Inc., Cambridge, Masschusetts, August 1979.

[14] P. R. Cohen. The pragmatics of referring and the modality of communication. *Computational Linguistics*, 10(2):97–146, April-June 1984.

[15] P. R. Cohen and H. J. Levesque. *Persistence, Intention, and Commitment.* Technical Report 415, Artificial Intelligence Center, SRI International, Menlo Park, California, February 1987. Also appears in *Proceedings of the 1986 Timberline Workshop on Planning and Practical Reasoning*, Morgan Kaufman Publishers, Inc. Los Altos, California.

[16] P. R. Cohen and H. J. Levesque. Speech acts and rationality. In *Proceedings of the Twenty-third Annual Meeting*, pages 49–59, Association for Computational Linguistics, Chicago, Illinois, July 1985.

[17] P. R. Cohen and H. J. Levesque. Speech acts and the recognition of shared plans. In *Proceedings of the Third Biennial Conference*, pages 263–271, Canadian Society for Computational Studies of Intelligence, Victoria, B. C., May 1980.

[18] P. R. Cohen and H. J. Levesque. Speech acts in a theory of rational interaction. 1987. In preparation.

[19] D. Davidson. Actions, reasons, and causes. In A. R. White, editor, *The Philosophy of Action*, Oxford University Press, 1968.

[20] A. I. Goldman. *A Theory of Human Action.* Princeton University Press, Princeton, New Jersey, 1970.

[21] D. Gordon and G. Lakoff. Conversational postulates. In *Papers from the Seventh Regional Meeting*, pages 63–84, Chicago Linguistic Society, 1971.

[22] H. P. Grice. Logic and conversation. In *Syntax and Semantics: Speech Acts*, Academic Press, New York, 1975.

[23] H. P. Grice. Meaning. *Philosophical Review*, 66:377–388, 1957.

[24] H. P. Grice. Utterer's meaning and intentions. *Philosophical Review*, 68(2):147–177, 1969.

[25] B. J. Grosz and C. L. Sidner. Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12(3):175–204, July-September 1986.

[26] B. J. Grosz and C. L. Sidner. Discourse structure and the proper treatment of interruptions. In *Proceedings of the Ninth International Joint Conference on Artificial Intelligence*, IJCAI, Los Angeles, California, August 1985.

[27] J. Y. Halpern and Y. O. Moses. A guide to the modal logics of knowledge and belief. In *Proceedings of the Ninth International Joint Conference on Artificial Intelligence*, IJCAI, Los Angeles, California, August 1985.

[28] D. Harel. *First-Order Dynamic Logic.* Springer-Verlag, New York City, New York, 1979.

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

[30] R. C. Moore. *Reasoning about Knowledge and Action.* Technical Note 191, Artificial Intelligence Center, SRI International, Menlo Park, California, October 1980.

[31] C. R. Perrault. An application of default logic to speech act theory. In P. R. Cohen, J. Morgan, and M. E. Pollack, editors, *Intentions in Communication*, M.I.T. Press, Cambridge, Massachusetts, 1987.

[32] C. R. Perrault and J. F. Allen. A plan-based analysis of indirect speech acts. *American Journal of Computational Linguistics*, 6(3):167–182, 1980.

[33] C. R. Perrault and P. R. Cohen. It's for your own good: a note on inaccurate reference. In *Elements of Discourse Understanding*, Cambridge University Press, Cambridge, Massachusetts, 1981. Also appears as Technical Report No. 4273, Bolt Beranek and Newman, Inc., Cambridge, Massachusetts, July, 1981.

[34] M. E. Pollack. A model of plan inference that distinguishes between the beliefs of actors and observers. In *Proceedings of the 24th Annual Meeting, Association for Computational Linguistics*, New York City, New York, 1986. Reprinted in: M. Georgeff and A. Lanksy, eds. *The 1986 Workshop on Reasoning about Actions and Plans*, Morgan Kaufmann Publishers, Los Altos, California.

[35] M. E. Pollack. Plans as complex mental attitudes. In P. R. Cohen, J. Morgan, and M. E. Pollack, editors, *The Role of Intentions and Plans in Communication and Discourse*, M.I.T. Press, Cambridge, Massachusetts, 1987.

[36] V. R. Pratt. *Six Lectures on Dynamic Logic.* Technical Report MIT/LCS/TM-117, Laboratory for Computer Science, MIT, Cambridge, Massachusetts, 1978.

[37] J. Sadock. *Toward a Linguistic Theory of Speech Acts.* Academic Press, New York, 1984.

[38] J. R. Searle. Indirect speech acts. In *Syntax and semantics, Speech acts*, Academic Press, New York, 1975.

[39] J. R. Searle. *Intentionality: An Essay in the Philosophy of Mind.* Cambridge University Press, New York City, New York, 1983.

[40] J. R. Searle. *Introduction*, pages 1–12. Oxford University Press, London, U. K., 1971.

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

[42] J. R. Searle and D. Vanderveken. *Foundations of Illocutionary Logic.* Cambridge Univ. Press, New York City, New York, 1985.

[43] F. Strawson. Intention and convention in speech acts. *The Philosophical Review*, v(lxxiii), 1964. Reprinted in *Logico-linguistic papers*, London: Methuen Co., 1971.

[44] T. Winograd and F. Flores. *Understanding Computers and Cognition: A New Foundation for Design.* Ablex Publishing Co., Norwood, New Jersey, 1986.