

# Task-Oriented Dialogue as a Consequence of Joint Activity\*

Philip R. Cohen  
Artificial Intelligence Center  
SRI International  
Menlo Park, CA, USA 94025  
pcohen@ai.sri.com

Hector J. Levesque<sup>†</sup>  
Dept. of Computer Science  
University of Toronto

José H. T. Nunes  
Dept. of Computer Science  
University of Toronto  
and

Sharon L. Oviatt  
Artificial Intelligence Center  
SRI International

## Abstract

We argue that current plan-based theories of discourse do not by themselves explain even simple task-oriented dialogues. The purpose of this paper is to show how a number of difficult-to-explain features of these dialogues follow from the *joint* or *team* nature of the underlying task. Specifically, the paper formally defines the concept of a joint intention and we argue that the conversants in a task-oriented dialogue jointly intend to accomplish the task. From this basis, the paper derives the goals underlying the pervasive use of confirmations, elaborations, requests for clarification, and temporal discourse markers observed in a recent experiment. We conclude with a discussion on generalizing the analysis presented here to characterize dialogue itself as a joint activity.

## 1 Introduction

Although it is a truism to say that dialogue is a joint endeavor, theories of communication, notably speech act theory, place most of the burden of communication on the speaker. Because definitions of illocutionary acts in many speech act theories (Cohen and Levesque 1990b; Searle 1969; Searle and Vanderveken 1985) require the speaker to be intending to convey his mental state, the speaker is committed to achieving understanding. But, so far as these theories go, the listener is not committed to anything at all. If the speaker fails in his communication, that is his problem. Nothing explains why listeners do not walk away in the middle of a dialogue, why they ask clarification questions, why they initiate repairs, or why they confirm.

Earlier work by ourselves and our colleagues (Allen and Perrault 1980; Cohen and Perrault 1979; Hobbs and Evans 1980; Perrault and Allen 1980) has proposed that dialogue is a special case of planned behavior, in which the actions to be planned are speech acts. These theories of dialogue contend that people plan their actions, either prospectively or reactively, to achieve their goals. Among the actions they plan are speech acts, leading them to issue utterances

in service of those goals of altering the addressee's mental states. (Appelt 1985; Cohen and Perrault 1979). Conversely, on hearing an utterance, a listener attempts to recognize what plan(s) the utterance serves (Allen and Perrault 1980; Litman and Allen 1990; Perrault and Allen 1980). Then, if helpfully disposed, he adopts goals to overcome obstacles in the speaker's plan. This latter phase of helpful goal adoption is the only way to date that these theories attribute new high-level goals to a listener that then lead to a response.

The goal of such plan-based theories is to derive linguistic phenomena from nonlinguistic analyses of rational, goal-directed behavior. As useful as this approach has been to the study of natural language dialogue, by itself it fails to explain fully the prevalent use of dialogue phenomena such as confirmations, requests for clarification, elaboration, and discourse or organizational markers (e.g., "Ok, the first thing"), etc. That is, they do not explain why these common properties of dialogue occur so regularly. Essentially, the reason for this failing is that such theories do not tell us when the goals underlying such speech acts would arise. Without such analyses, plan-based theories cannot determine when to plan such speech acts.

Some of these concerns have begun to be addressed with the separation of domain plans from discourse plans (Litman and Allen 1990), and attentional, intentional, and linguistic structures (Grosz and Sidner 1986). However, what is lacking from even these plan-based approaches is a *competence* theory of dialogue that provides independent specifications for the correctness of processing models and predicts discourse phenomena. By "predict" and "explain" we have in mind a derivation from independently motivated principles. The method of derivation used here is that of logical consequence, as one's theories are forced to be clear and reproducible. As for independent motivation, our principles will be derived from an analysis of teamwork, and will be applied to situations that result in dialogue. Specifically, this paper shows how many of the properties of discourse mentioned above follow from the adoption of joint intentions. For technical reasons that will be explained later, we restrict our attention to task-oriented dialogues, dialogues occurring in a setting in which one party instructs another in the performing a task.

\*This paper was supported by a contract from ATR International to SRI International, and by a gift from the Systems Development Foundation. The Toronto authors were supported in part by a grant from the Natural Sciences and Engineering Research Council of Canada. Our thanks go to Marco Ponzi for valuable comments.

<sup>†</sup>Fellow of the Canadian Institute for Advanced Research

## 2 Joint action and dialogue

Intuitively, dialogue is a paradigm case of joint action. Each of the conversants tries to make himself or herself understood, and each is committed to understanding the other and to the other's being successful in achieving his/her goals. Joint activity, we argue, rests on the agents' having joint commitments, and those joint commitments to engage in dialogue lead to elaborations, repairs, clarifications, confirmations and other signals of understanding, requests for confirmation, and the like. This linguistic behavior needed to accomplish a joint activity forms a scaffold upon which the conversants achieve their collective and individual goals. The foundation of this scaffold, and frequently of the edifice under construction, is the theory of joint activity.

Joint action, or doing something together as a team, is more than merely coordinating individual agents' actions: stop signs coordinate drivers, thereby regulating their actions, but do not make drivers act as a team. Being helpful to other agents, even recognizing that one is intended to be helpful, also does not fully constitute team behavior — for example, one might know by a turn signal that driver *A* in the next lane wants *B* to let him into *B*'s lane. Even if *B* does so, *A* and *B* are not therefore driving as a team.

But when *A* and *B* are driving somewhere together, each party is committed to the other's success, and those other-directed commitments constrain each agent's own adoption of actions. For example, if *A* and *B* are jointly committed to *B*'s following *A* until *B* knows where to go, we would predict these actions: there will be a signal when it is time to get started, *A* and *B* will endeavor to keep each other within sight and not pull over privately, *A* will not take actions that he knows would render *B*'s intentions impossible to achieve, *B* will signal when he knows his way, and without such a signal, *A* will still assume *B* is following. Of course, *A* and *B* do not need to explicitly agree to these actions; these should be consequences of what it means to act together. As we shall see, analogous actions occur in task-oriented dialogue.

Earlier exponents of the position that dialogue is a kind of joint action include Clark and Wilkes-Gibbs (1986) and Grosz and Sidner (1990). Clark and Wilkes-Gibbs point out that referring needs to be viewed as collaborative, in that speakers present descriptions and listeners accept them as adequate, and seek repairs if not. They argue that the key goal is for the conversants to attain a state of mutual belief that the description is adequate for present purposes. However, they do not derive the conversants' goals for attaining mutual belief from other principles. Grosz and Sidner also argue for the need to view dialogue as a collaborative activity. They propose a definition of joint plans and argue that such mental states are needed to support dialogue. We agree with many of their observations, but find their analysis of joint plans deficient in ways that preclude the kinds of derivations presented here.

In a companion paper (Levesque *et al.* 1990a), we define and explain the concept of joint commitment parallel to our treatment of individual commitment (Cohen and Levesque 1990a). In that paper, we show how agents' adoption of joint commitments entails their having individual commitments to do their parts. Here, we apply the theory to derive the goals underlying the use of linguistic devices.

In the remainder of the paper, we first present the data to be explained. Next, a brief exposition of our theory of joint intention and commitment is given. Then, a model of the task-oriented dialogue situation as a joint activity is provided, and the theory is applied to explain the data. Finally, we discuss extending the theory to handle dialogue more generally as a joint activity.

## 3 The data

In a recent report (Oviatt and Cohen 1988), two of us analyzed how the structure of ten simple expert-apprentice task-oriented dialogues for assembling a toy water pump depended on the modality of communication, where the modalities included telephone and audiotape recording. Among the findings were the following:<sup>2</sup>

- In the telephone dialogues, successful satisfaction of each expert request was nearly always confirmed with utterances such as "Done." Confirmations of referent identification (including many so-called "back-channel" utterances, such as "mm-hmm") were also frequent. These two categories, confirmation of successful action and of successful referent identification, accounted for 89% of all confirmations present in the telephone dialogues. When confirmations were slow or absent in the telephone mode, experts often explicitly requested them. Overall, 18% of the verbal interaction in telephone mode was spent eliciting and issuing confirmations, with an average rate of one confirmation every 5.6 seconds.
- In both modalities, discourse or organizational markers such as "next," "first thing," and "ok, now" initiated 96%-99% of the discourse segments pertaining to the assembly of a specific pump part (hereafter termed "discourse assembly segments").
- There were an average of 21 and 35 elaborations (including referential and action elaborations) in each telephone and audiotape dialogue, respectively.
- Telephone dialogues averaged 5.4 clarification subdialogues each.

Any adequate theory of dialogue should be able to explain these phenomena. In particular, why do these speech acts occur, and occur as often as they do? Consider confirmations, either of successful completion of a requested assembly action, or of successful identification of a pump part that has been referred to. No analysis of requests provided by Searle (1969; 1985) or by computational linguists (Allen and Perrault 1980; Appelt 1985; Cohen and Levesque 1990b; Cohen and Perrault 1979; Litman and Allen 1990) ourselves included, predicts that a listener will inform the speaker when the requested action is complete. The only explanatory recourse for plan-based theories is to rely on the listener's plan-recognition processing to discover a blocked goal that must be overcome for the speaker's overall plan to succeed. But, there need not be such a more general plan, in which case there would be no obstacles to overcome. Nonetheless, one would expect to find the confirmations. Essentially, we ask where did those goals come from?

One avenue of discourse research has been to investigate discourse structure in terms of abstract relations among segments of text. In this vein, "discourse markers" have been analyzed by many researchers, including, R. Cohen (1987), Grosz and Sidner (1986), Hobbs (1979), Litman and Allen (1990), Hirschberg and Litman (1987), Reichman (1981), and by Schiffrin (1987), as providing textual or intonational "clues" about the evolving discourse structure. Although some of these authors give mental state analyses for speech acts, they do not provide mental state semantics for the markers themselves, nor do they *derive* within their theories why those markers should occur, other than to say that speakers must signal how their discourse is being structured. Rather, most of these authors are concerned

<sup>2</sup> Definitions of all terms used below, the coding scheme, and interrater reliabilities can be found in (Oviatt and Cohen 1988).

with a system's or listener's using the markers to recognize the structure of the discourse, often to aid in analyzing referential phenomena. While not denying the computational utility of viewing the markers as signaling discourse segment relationships, one result of this paper is to explain how the *discourse intentions* (Grosz and Sidner 1986; Litman and Allen 1990) underlying many of the markers found in this study arise as consequences of joint activity.

We now proceed to describe formally what is meant here by joint commitments and joint intentions. Then, we characterize the task-oriented dialogue situation in terms of these concepts, and show how the analysis predicts the speakers' goals underlying the data.

## 4 The formalism

Our account of joint commitment and intention (Cohen and Levesque 1990a; Levesque *et al.* 1990a) is formulated in a modal language of belief, goal, action, and time. The language we will use has the usual connectives of a first-order language with equality, as well as operators for the propositional attitudes and for talking about sequences of events:  $(BEL\ x\ p)$  and  $(GOAL\ x\ p)$  say that  $x$  has  $p$  as a belief and goal respectively;  $(MB\ x\ y\ p)$  says that  $x$  and  $y$  mutually believe that  $p$  holds;  $(AGT\ x_1\ \dots\ x_n\ e)$  says that  $x_1\ \dots\ x_n$  are the only agents for the sequence of events  $e$ ;  $e_1 \leq e_2$  says that  $e_1$  is an initial subsequence of  $e_2$ ; and finally,  $(HAPPENED\ a)$ ,  $(HAPPENING\ a)$ , and  $(HAPPENS\ a)$  say that a sequence of events describable by an action expression  $a$  has just happened, is happening now, or will happen next, respectively. An action expression here is built from variables ranging over sequences of events using the constructs of dynamic logic:  $a;b$  is action composition;  $a|b$  is nondeterministic choice;  $a||b$  is concurrent occurrence of  $a$  and  $b$ ;  $p?$  is a test action; and finally,  $a^*$  is repetition. The usual programming constructs like IF/THEN actions and WHILE loops can easily be formed from these.<sup>3</sup> For remarks on the semantics of this language, see other papers of ours (Cohen and Levesque 1990a; Levesque *et al.* 1990a).

We will use the following syntactic abbreviations:

### Actions:

$$\begin{aligned} (DONE\ x_1\ \dots\ x_n\ a) &\stackrel{\text{def}}{=} (HAPPENED\ a) \wedge (AGT\ x_1\ \dots\ x_n\ a). \\ (DOING\ x_1\ \dots\ x_n\ a) &\stackrel{\text{def}}{=} (HAPPENING\ a) \wedge (AGT\ x_1\ \dots\ x_n\ a). \\ (DOES\ x_1\ \dots\ x_n\ a) &\stackrel{\text{def}}{=} (HAPPENS\ a) \wedge (AGT\ x_1\ \dots\ x_n\ a). \end{aligned}$$

**Eventually:**  $\Diamond p \stackrel{\text{def}}{=} \exists e (HAPPENS\ e; p?)$ .

There is something that happens, including the empty sequence of events, after which  $p$  holds, *i.e.*,  $p$  is true at some point in the future.

**Earlier:**  $(EARLIER\ p) \stackrel{\text{def}}{=} \neg p \wedge \exists e (HAPPENED\ p?; e)$ .

**Always:**  $\Box p \stackrel{\text{def}}{=} \neg \Diamond \neg p$ .

$\Box p$  means that  $p$  is true throughout the course of events from now on.

**Until:**  $(UNTIL\ p\ q) \stackrel{\text{def}}{=} \forall c (HAPPENS\ c; \neg q?) \supset \exists a (a \leq c) \wedge (HAPPENS\ a; p?)$ . Until the wff  $p$  is true, the wff  $q$  will remain true.

**Knowledge:**  $(KNOW\ x\ p) \stackrel{\text{def}}{=} p \wedge (BEL\ x\ p)$

<sup>3</sup>Test actions occur frequently in our analysis, yet are potentially confusing. The expression  $p?; a$  should be read as "action  $a$  with  $p$  holding initially," and analogously for  $a; p?$ . Note specifically that an agent can perform these without ever knowing the truth value of  $p$ , but if he intends to achieve such an action, he will eventually know the truth value of  $p$ .

**Mutual Knowledge:**  $(MK\ x\ y\ p) \stackrel{\text{def}}{=} (BEL\ x\ (MB\ x\ y\ p)) \wedge (BEL\ y\ (MB\ y\ x\ p))$

Based on these primitives, a notion of individual commitment called PGOAL, for persistent goal, has been defined (Cohen and Levesque 1990a) that describes an agent as being committed to  $p$  if he knows that he will keep his goal that  $p$  at least until he believes it is true, is impossible, or is irrelevant. More formally,<sup>4</sup>

**Definition 1**  $(PGOAL\ x\ p\ q) \stackrel{\text{def}}{=} (BEL\ x\ \neg p) \wedge (GOAL\ x\ \Diamond p) \wedge (KNOW\ x\ (UNTIL\ [(BEL\ x\ p) \vee (BEL\ x\ \Box \neg p) \vee (BEL\ x\ \neg q)] (GOAL\ x\ \Diamond p)))$

The important points to observe about individual commitment are that: Once adopted, an agent cannot drop them freely, other commitments need to be consistent with them, and agents will replan to achieve them should initial attempts fail. Condition  $q$  is an "escape clause" (which we will frequently omit for brevity), against which the agent has relativized his persistent goal. Should the agent come to believe it is false, he can drop the commitment. Frequently, the escape clause will encode the network of reasons why the agent has adopted the commitment. For example, with it we can make a commitment a subgoal, either to the agent's own supergoal, or to a (believed) goal of another agent.

Individual intention to do action  $a$  is defined as a PGOAL to do a thinking one is doing it.

**Definition 2**  $(INTEND_1\ x\ a\ q) \stackrel{\text{def}}{=} (PGOAL\ x\ (DONE\ x\ (UNTIL\ (DONE\ x\ a) (BEL\ x\ (DOING\ x\ a))))?; a)\ q)$

Thus, an intention is a commitment to do an action deliberately, *i.e.*, knowing what one is doing. Intentions inherit all the properties of commitments, and also have those that emerge from the interaction of belief and action (Cohen and Levesque 1990a).

## 5 Joint commitments and intentions

In another paper (Levesque *et al.* 1990a), we present, justify, and draw consequences from definitions of joint intention and commitment, parallel to our definitions of individual intention and commitment. The definitions are motivated by nonlinguistic examples, *e.g.*, driving in a convoy, and by a principle of making minimal changes to the analysis of individual intentions and commitments. Here, we can only state the definitions and show how they lead to dialogue.

**Definition 3**  $(JPG\ x\ y\ p\ q) \stackrel{\text{def}}{=} (MB\ x\ y\ \neg p) \wedge (MB\ x\ y\ [(GOAL\ x\ \Diamond p) \wedge (GOAL\ y\ \Diamond p)]) \wedge (MK\ x\ y\ [UNTIL\ [(MB\ x\ y\ p) \vee (MB\ x\ y\ \Box \neg p) \vee (MB\ x\ y\ \neg q)] (WGM\ x\ y\ p\ q)]), \text{ where}$

$(WGM\ x\ y\ p\ q) \stackrel{\text{def}}{=} (MB\ x\ y\ (WG\ x\ y\ p\ q) \wedge (WG\ y\ x\ p\ q)), \text{ where}$

$(WG\ x\ y\ p\ q) \stackrel{\text{def}}{=} [\neg (BEL\ x\ p) \wedge (GOAL\ x\ \Diamond p)] \vee [(BEL\ x\ p) \wedge (GOAL\ x\ \Diamond (MB\ x\ y\ p))] \vee [(BEL\ x\ \Box \neg p) \wedge (GOAL\ x\ \Diamond (MB\ x\ y\ \Box \neg p))] \vee [(BEL\ x\ \neg q) \wedge (GOAL\ x\ \Diamond (MB\ x\ y\ \neg q))]$

<sup>4</sup>This definition differs slightly from that presented in our earlier work (Cohen and Levesque 1990a), but that difference is irrelevant here.

A joint persistent goal (JPG) to achieve  $p$  relative to  $q$  means that the agents mutually believe  $p$  is false, they mutually believe that they each want it to be true, and they mutually know they will keep the "weak mutual goal" WMG at least until they mutually believe  $p$  holds, is impossible, or irrelevant. An agent's having a weak mutual goal with another agent consists of a mutual belief that each agent has a "weak goal," which itself consists of an agent's either having the goal to achieve  $p$ ; or, if he believes it to be true, the goal of achieving mutual belief that it is true; or, if he believes it to be impossible, the goal to make it mutually believed that it is impossible; or, if he believes it to be irrelevant ( $\neg q$ ), the goal to make it mutually believed that it is irrelevant. The concept of a weak goal is needed because once two parties agree to do something, one cannot be absolutely sure the agreement is still on. All one can know for certain is that either the agreement is still on, or the other party is trying to communicate that it is off. Ruled out by this analysis is the case in which the other party drops the agreement privately.

By analogy with individual commitments, joint commitments need to satisfy joint versions of Bratman's (1987) functional roles for intention, namely they should lead groups to do joint planning, they should screen options incompatible with the group's success, and they should persist and be tracked until termination. A more comprehensive paper (Levesque *et al.* 1990b) discusses how these properties arise from the definition. Here, it is important to notice that an agent can be jointly and individually committed to the action(s) of another agent.

### 5.1 Joint Intention

We analyze two agents' jointly intending to perform a specific action as follows:

**Definition 4** ( $JI_1 \times y \ a \ q$ )<sup>def</sup>  
 $(JPG \times y \ \exists t \ (DONE \times y \ t?:a) \ (MB \times y \ (DOING \times y \ t?:a)))?:t?:a \ q$

That is, jointly intending to do an action  $a$  is a joint commitment to do that action while both members mutually believe they are doing it starting at  $t$ .<sup>5</sup>

Many interesting properties can be shown to follow from having joint intentions. For example, joint commitments and intentions lead agents to have individual commitments and intentions

**Theorem 1** *Joint and individual commitments and intentions*

- (A)  $\models (JPG \times y \ p \ q) \supset (PGOAL \times p \ q)$   
 (B)  $\models (JI_1 \times y \ a \ q) \supset (INTEND_1 \times a \ q)$ , provided  $x$  knows he is the only agent of  $a$ .

Notice that by this theorem, two agents can jointly intend for one of them to act. That agent has an individual intention, and the other has a persistent goal. The proof of part A is given in (Levesque *et al.* 1990a), and that of part B follows immediately from the definitions of joint intention and mutual belief. This theorem, and its corollaries, explain one of the puzzles of joint action noted by Searle (1990), namely, how joint intention results in individual intentions. A similar property is stated as part of the *definition* of joint plans by Grosz and Sidner (1990) and of "we-intentions" by Tuomela and Miller (1988).

Often, we will want to consider individual and joint intentions to do something to *achieve* a proposition  $p$ , where the agent(s) need not know in advance what he (they) will do. A slightly modified version of individual intention to achieve  $p$ ,  $INTEND_2$ , from (Cohen and Levesque 1990a) is:

<sup>5</sup>The important temporal relationship here is that the agents mutually know that they are starting together, not what time it is.

**Definition 5** ( $INTEND_2 \times p \ q$ )<sup>def</sup>  
 $(PGOAL \times \exists e, t \ (DONE \times \ [UNTIL \ p \ (BEL \times \exists e' \ (DOING \times t?:e';p?))?:t?:e;p?] \ q)$

That is, the agent is committed to doing some event  $e$  at some time  $t$  to achieve  $p$ , all the while believing that he is in the process of doing something ( $e'$ ) to achieve  $p$ .

Below, we give an analogous definition for jointly intending to achieve  $p$ .

**Definition 6** ( $JI_2 \times y \ p \ q$ )<sup>def</sup>  
 $(JPG \times y \ \exists e, t \ (DONE \times y \ [UNTIL \ p \ (MB \times y \ \exists e' \ (DOING \times y \ t?:e';p?))?:t?:e;p?] \ q)$

Two agents,  $x$  and  $y$  jointly intend to achieve  $p$  if the two agents are jointly committed to doing some event  $e$  starting at some time  $t$  to achieve  $p$ , not necessarily knowing at the time they adopt the joint intention precisely what they will do. Part of their joint commitment is that they mutually believe that they are achieving it (though not necessarily with  $e$ ) while they are doing so, starting together at time  $t$ .

In summary, joint commitment and intention are natural extensions of our analysis of individual commitment and intention, and satisfy an extensive set of adequacy criteria based on nonlinguistic properties. We now proceed to show how this analysis predicts linguistic phenomena found in our study.

## 6 Communication

First, consider confirmations. Both parties are committed to the apprentice's confirming the successful completion of assembly actions. We can now show the origin of this goal by proving a more general theorem to the effect that each party is committed to achieving a mutual belief of success.

**Theorem 2**  $\models (JPG \times y \ p \ q) \wedge C \supset$   
 $(UNTIL \ [((MB \times y \ p) \vee (MB \times y \ \neg p) \vee (MB \times y \ \neg q)) \ ((BEL \times p \wedge \neg(MB \times y \ p)) \supset (PGOAL \times (MB \times y \ p)))])$

That is, if  $x$  and  $y$  are jointly committed to  $p$ , and some condition  $C$  holds (discussed below), then until the agents mutually believe that  $p$  is satisfied, or is impossible, if one agent, say  $x$  comes to believe  $p$  and does not yet mutually believe with his partner that  $p$  holds, then he has the persistent goal to make  $p$  mutually believed.

*Proof sketch:* Assume that  $x$ , but not  $y$ , comes to believe  $p$  is satisfied (impossible).<sup>6</sup>  $C$  is some condition that is sufficient to guarantee that  $x$  will not change his mind about the satisfaction (resp. impossibility) of  $p$ . The simplest and strongest such condition for satisfaction, if  $p$  is time-limited (i.e.,  $p$  must be true at some time  $t$ ), is  $\Box[(BEL \times p) \supset \Box(BEL \times p)]$ , but others are possible. The  $JPG$  to achieve  $p$  becomes false, because  $\neg(MB \times y \ \neg p)$ . However,  $x$  cannot drop his weak mutual WMG goal until  $(MB \times y \ p)$  holds (or  $(MB \times y \ \neg p)$ ). But, this means he cannot drop his weak goal (WG) until then either. The operative disjunction in that weak goal is his believing  $p$  and having the goal that eventually  $(MB \times y \ p)$  holds. So, one can see immediately that  $(PGOAL \times (MB \times y \ p))$  holds.

Given our speech act theory (Cohen and Levesque 1990b), there are many possible ways for the agents to attain the relevant states of mutual belief. For example, one can show that a sequence of two assertions that the requested action has been done, where neither party doubts the sincerity of the other, will discharge the residual individual commitments to achieve  $(MB \times y \ p)$  that are entailed by a joint commitment.<sup>7</sup> A question by one party (e.g., "Done?"), serving as a request for confirmation, followed by an assertion, would do as well.

<sup>6</sup>We ignore the case of irrelevance as it is similar.

<sup>7</sup>For a proof, please see (Levesque *et al.* 1990b).

## 6.1 Initiating a Jointly Intended Action

Not only must joint intentions and commitments terminate with mutual belief, they must commence that way as well. That is, there must be at least a state of  $(MB \times y \text{ (DOES } \times y \text{ a)})$ , where  $a$  is the jointly intended action. In other words, the agents must mutually believe they are about to start. The following theorem shows this:

**Theorem 3** *Starting a jointly intended action*  
 $\models (Jl_1 \times y \text{ a}) \wedge \neg[\text{EARLIER (MB } \times y \text{ (DOES } \times y \text{ a}) )]$   
 $\supset (PGOAL \times \text{ (MB } \times y \text{ (DOES } \times y \text{ a)}) [Jl_1 \times y \text{ a}] \wedge$   
 $(PGOAL \text{ y (MB } \times y \text{ (DOES } \times y \text{ a)}) [Jl_1 \times y \text{ a}])$

That is, if the agents have adopted a joint intention to achieve action  $a$ , then they are committed to its being mutually believed that they are starting, provided they still have the joint intention and they have yet to mutually believe they have started. For reasons of space, we omit the proof here (but see (Levesque *et al.* 1990b)).

The seemingly required use of temporal discourse markers, such as "next," before each discourse assembly segment should now be clear — they are needed to synchronize before initiating the jointly intended actions. This usage accords with various analyses (Hirschberg and Litman 1987; Schiffrin 1987) of the marker usage of "now," as signalling the speaker's progression from one topic to another.

## 6.2 Modeling Task-Oriented Dialogue

We have argued informally elsewhere (Oviatt and Cohen 1988) that in our telephone and audiotape conditions, the expert and apprentice jointly intended to perform the task. One way to model that task-oriented dialogue situation is to say that the expert and apprentice had the following joint intentions:

- (1)  $(Jl_2 \text{ exp appr (Assembled Pump)})$ , which leads to
- (2)  $\forall p \in \{\text{Parts}\}$   
 $(Jl_2 \text{ exp appr } \exists e' \text{ (AGT } e' \text{ appr)}) \wedge$   
 $([\text{KNOW appr (GOAL exp (DOES appr } e'))]?) \wedge e' : (\text{Assembled } p?)$

The second formula says that for every part, the conversants had a joint intention for the apprentice to do some action  $e'$  thinking that  $e'$  is what the expert wants the apprentice to do next, after which the part would be assembled.<sup>8</sup>

The analysis of joint intention predicts the presence of the *goals* or *discourse intentions* achieved by the speech acts noted below.

- For each part, the expert will need to get the apprentice to know what he wants him to do next, and to get the apprentice to then do so. *Requests* are the typical means used to achieve these goals, and elsewhere (Cohen and Levesque 1990b), we have analyzed requests as attempts to do so. Another possibility, though one rarely found in the data, is that the apprentice can initiate a discourse assembly segment by asking a question about what action is to be done next. Finally, this formulation allows the apprentice to assemble something before the expert requests it, but he must convince himself that the action is desired next, and must confirm its success (see below).

<sup>8</sup>Notice that through the definition of  $Jl_2$ , we have been careful not to quantify the apprentice's assembly event ( $e'$ ) into the mutual belief of what the agents are doing. Were we to do so, the apprentice would be required to know (in fact, to mutually believe) in advance what is to be requested next.

- If after one attempt, the expert does not think he has achieved the goal of getting the apprentice to know what should be done next, the persistence of the expert's commitment results in an *elaboration* as a second (or subsequent) attempt to achieve the goal.
- If the apprentice does not understand what assembly action the expert wants him to do next, he has a persistent goal to acquire that knowledge. Thus, *clarification questions* conditionally follow from the joint intention.
- For each part, the participants are committed to there being a state of mutual belief attained that the part has been assembled (see Theorem 2). When the conversants are not copresent, this state may be achieved with two utterances — an assertion of completion by the apprentice (e.g., "Done" or "OK") which is one form of *confirmation* and an acknowledgement by the expert.
- If the expert does not receive the action confirmation, he does not merely move on to another subtask. Rather, his commitment to attaining mutual belief of success persists (see Theorem 2), resulting in *requests for confirmation*.
- The participants need to attain the mutual belief that a new jointly intended action is beginning (see Theorem 3). This happens through the use of *temporal discourse markers* before each discourse assembly segment.
- If one assumes that an agent's knowing what constitutes given event (type) entails his knowing what all the arguments to that event (type) are (Moore 1985), then this analysis predicts the goals underlying the *referential elaborations* and *confirmations of successful identification* so common in the transcripts, and also described by Clark and Wilkes-Gibbs (1986). That is, in virtue of the partners' joint intention for the apprentice to perform an assembly action, the agents are jointly committed to the apprentice's knowing what pump parts that action involves. Therefore, all the aforementioned speech acts apply to part identification just as they apply to assembly actions.

The important point to notice here is not that the discourse intention behind any one of these phenomena is predicted, but that one analysis predicts all of them. Future research will attempt to derive other phenomena as well. For example, during a jointly intended act, the parties are supposed to mutually believe they are performing it. Hence, once they have synchronized, signals are needed in order to keep the mutual belief from dissipating over time.

## 7 Concluding remarks: dialogue as a joint activity

In this paper, we have considered only task-oriented dialogues. A substantive promissory note of this approach is to view dialogue more generally as a joint activity, one that is appropriately initiated, monitored, and closed, and is robust against miscommunication. There are two reasons for considering only task-oriented dialogues in detail here. First, though these dialogues are very simple in structure, there is as yet no satisfying account of them. Much can be gained by sharpening one's tools on the simple cases first.

Second, there is a technical difficulty looming. We were able to derive numerous predictions about discourse phenomena from the single assumption that the two conversants had a joint intention for the apprentice to assemble a part once he knew what assembly action the expert wanted him to do

next. That is, the apprentice was committed to doing *what-ever* action the speaker wanted him to do. Formally, this involved only quantifying over events. In a more general setting, where we would want to say that the conversants were jointly committed to understanding one another, we would need to quantify over propositions. That is, we need to say that the parties are jointly committed to making mutually believed that the speaker wants the addressee to know that he (the speaker) is in a certain mental state, described by a proposition  $\Psi$ . As the listener does not know what the speaker is about to say, to describe the listener's view of that commitment, we would need to quantify over the  $\Psi$ 's. This we cannot do with our possible-worlds approach, but look forward to a situation theory (Barwise and Perry 1983) to provide the technical apparatus.

Finally, we believe the properties of dialogue discussed here are not simply a result of the interaction of plan generators and recognizers working in synchrony and harmony, as plan-based theories propose. Rather, what Clark and Wilkes-Gibbs (1986), Grosz and Sidner (1990), and we are suggesting is that *both* parties in a dialogue are responsible for sustaining it. Participating in a dialogue requires the conversants to have at least a joint commitment, if not a joint intention, to make themselves understood. The key question to be answered is how to formalize such general commitments precisely, and to show how they predict the fine-grained synchrony so apparent in ordinary conversation.

## References

- [Allen and Perrault 1980] J. F. Allen and C. R. Perrault. Analyzing intention in dialogues. *Artificial Intelligence*, 15(3):143-178, 1980.
- [Appelt 1985] D. Appelt. *Planning English Sentences*. Cambridge University Press, Cambridge, U. K., 1985.
- [Barwise and Perry 1983] J. Barwise and J. Perry. *Situations and Attitudes*. MIT Press, Cambridge, Massachusetts, 1983.
- [Bratman 1987] M. Bratman. *Intentions, Plans, and Practical Reason*. Harvard University Press, 1987.
- [Clark and Wilkes-Gibbs 1986] H. H. Clark and D. Wilkes-Gibbs. Referring as a collaborative process. *Cognition*, 22:1-39, 1986. Reprinted in: *Intentions in Communication*, Cohen, P. R., Morgan, J., and Pollack, M. E., editors. MIT Press, Cambridge, Massachusetts, 1990.
- [Cohen and Levesque 1990a] P. R. Cohen and H. J. Levesque. Intention is choice with commitment. *Artificial Intelligence*, 42(3), 1990.
- [Cohen and Levesque 1990b] P. R. Cohen and H. J. Levesque. Rational interaction as the basis for communication. In P. R. Cohen, J. Morgan, and M. E. Pollack, editors, *Intentions in Communication*. M.I.T. Press, Cambridge, Massachusetts, 1990.
- [Cohen and Perrault 1979] P. R. Cohen and C. R. Perrault. Elements of a plan-based theory of speech acts. *Cognitive Science*, 3(3):177-212, 1979.
- [Cohen 1987] R. Cohen. Analyzing the structure of argumentative discourse. *Computational Linguistics*, 13(1-2):11-24, 1987.
- [Grosz and Sidner 1986] B. J. Grosz and C. L. Sidner. Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12(3):175-204, July-September 1986.
- [Grosz and Sidner 1990] B. Grosz and C. Sidner. Plans for discourse. In P. R. Cohen, J. Morgan, and M. E. Pollack, editors, *Intentions in Communication*. M.I.T. Press, Cambridge, Massachusetts, 1990.
- [Hirschberg and Litman 1987] J. Hirschberg and D. Litman. Now let's talk about *now*: Identifying cue phrases intonationally. In *Proceeding of the 25th Annual Meeting of the Association for Computational Linguistics*, pages 163-171, Stanford, California, June 1987.
- [Hobbs and Evans 1980] J. Hobbs and D. Evans. Conversation as planned behavior. *Cognitive Science*, 4(4):349-377, 1980.
- [Hobbs 1979] J. Hobbs. Coherence and coreference. *Cognitive Science*, 3(1):67-90, 1979.
- [Levesque et al. 1990a] H. J. Levesque, P. R. Cohen, and J. Nunes. On acting together. In *Proceedings of AAAI-90*, San Mateo, California, July 1990. Morgan Kaufmann Publishers, Inc.
- [Levesque et al. 1990b] H. J. Levesque, P. R. Cohen, and J. Nunes. On acting together: Joint intentions for intelligent agents. Unpublished ms., 1990.
- [Litman and Allen 1990] D. J. Litman and J. F. Allen. Discourse processing and commonsense plans. In P. R. Cohen, J. Morgan, and M. E. Pollack, editors, *Intentions in Communication*, pages 365-388. M.I.T. Press, Cambridge, Massachusetts, 1990.
- [Moore 1985] R. C. Moore. A formal theory of knowledge and action. In *Formal Theories of the Commonsense World*, chapter 9, pages 319-358. Ablex Publishing Co., Norwood, New Jersey, 1985.
- [Oviatt and Cohen 1988] S. L. Oviatt and P. R. Cohen. Discourse structure and performance efficiency in interactive and noninteractive spoken modalities. Technical Report 454, Artificial Intelligence Center, SRI International, Menlo Park, California, 1988. To appear in *Computer Speech and Language*.
- [Perrault and Allen 1980] 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.
- [Reichman 1981] R. Reichman. *Plain-speaking: A theory and grammar of spontaneous discourse*. PhD thesis, Department of Computer Science, Harvard University, Cambridge, Massachusetts, 1981.
- [Schiffrrin 1987] D. Schiffrrin. *Discourse Markers*. Cambridge University Press, Cambridge, U. K., 1987.
- [Searle and Vanderveken 1985] J. R. Searle and D. Vanderveken. *Foundations of Illocutionary Logic*. Cambridge Univ. Press, New York City, New York, 1985.
- [Searle 1969] J. R. Searle. *Speech acts: An essay in the philosophy of language*. Cambridge University Press, Cambridge, 1969.
- [Searle 1990] J. R. Searle. Collective intentionality. In P. R. Cohen, J. Morgan, and M. E. Pollack, editors, *Intentions in Communication*. M.I.T. Press, Cambridge, Massachusetts, 1990.
- [Tuomela and Miller 1988] R. Tuomela and K. Miller. We-intentions. *Philosophical Studies*, 53:367-389, 1988.