Confirmations and Joint Action*

Philip R. Cohen
Artificial Intelligence Center
SRI International
Menlo Park, CA, USA 94025

Hector J. Levesque[t]
Dept. of Computer Science
University of Toronto

Abstract

We argue that current plan-based theories of discourse do not by themselves explain prevalent phenomena in even simple task-oriented dialogues. The purpose of this paper is to show how one difhcult-to-explain feature of these dialogues, confirmations, follows from the *joint* or team nature of the underlying task. Specifically, we review 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, we derive the goals underlying the pervasive use of confirmations 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

This paper is concerned with analyzing features of communication that arise during joint or team activities. Specifically, we discuss the origin of two types of confirmations: confirmations of successful fulfillment of requests, and of referent identification. We show how the same underlying principles, namely the joint nature of the partners' activity, gives rise to the goals underlying both types of speech acts. To make this precise, we examine data from a study of task-oriented dialogues, provide a formal analysis of joint action, describe the subjects' task in the formalism, and then derive the goals that we claim give rise to the confirmations.

Both types of confirmation, as well as requests for them, occur repeatedly in the task-oriented dialogues for assembling a toy water pump studied in [4, 12]. First, successful satisfaction of each request from an expert to an apprentice was nearly always confirmed with utterances such as "Got it," "OK," or "Done." For example,[1]

*Thi8 paper was supported by a contract from ATR International to SRI International, by a gift from the Systems Development Foundation, and by a grant from the Natural Sciences and Engineering Research Council of Canada. The second author also wishes to thank the Department of Computer Science and the Center for the Study of Language and Information at Stanford University where he was a visitor during the preparation of this paper.

+ Fellow of the Canadian Institute for Advanced Research

Dialogue fragments are quoted verbatim from transcriptions from the study, but with emphasis added for clarity.

Exp: And attach the pink thing so it covers the hole in the middle.
Appr: (pause) *Got it.* One way-valve. We're all set.

The confirmation clearly indicates that the apprentice thinks he has successfully assembled the relevant part, and not simply that the request was understood or would be complied with. Similarly, confirmations of referential understanding, including many so-called "back-channel" utterances, such as "mm-hmm" or "Yeah", were also frequent. For example,

Exp: Okay, I want you to take the largest tube, or actually
it's the largest piece of anything,
that has two openings on the side -
Appr: *Yeah*
Exp: - and threads on the bottom.
Appr: *Yeah.*

These two categories, confirmation of successful action and of successful referent identification, accounted for 89% of all confirmations in the telephone dialogues. When confirmations were slow or absent in these dialogues, experts often explicitly requested them. For example, requests for both types of confirmation can be found in the following fragment:

Exp: And stick it on the en-onto the uh spout coming out the side.
You see that?
Appr: Yeah, okay
Exp: *You got that on, okay?*
Appr: Yeah.

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.

Clearly, confirmations and requests for them are such a crucial component of dialogue success, in this task and many others, that any adequate theory of dialogue should be able to explain why and where they should occur. However, no current plan-based theories of dialogue or speech act theories do so.[2] Essentially, the reason for

[2]Because space precludes an extensive discussion, we will assume the reader is familiar with those theories. If not, please see, for example, [1, 7, 9, 11].

this failure is that such theories do not tell us when the goals underlying such speech acts would arise.

To be more specific, how would plan-based theories of dialogue fashioned after the Allen and Perrault model [1] attempt to account for the apprentice's confirmations and the expert's requests for them? Essentially, there are three possible routes: First, the apprentice's goal of attaining mutual belief of successful action (which would lead to one kind of confirmation) could be added to the semantics of a request. Thus, in addition to conveying what action the speaker wants the addressee to perform, a request would also convey that the speaker wants the outcome of the requested action to become mutually believed. Since these plan-based theories required illocutionary act recognition [1, 11], if a request were recognized as part of the expert's plan, so too would his goal of attaining a mutual belief (or perhaps, just a belief) that the requested action has been performed. Then, by helpful goal adoption, the goal of attaining mutual belief of successful action would be adopted by the apprentice. This augmentation of the meaning of requesting might handle the problem, but no argument has been given on independent grounds for doing so.

Second, in the course of attempting to recognize the expert's higher-level plan, such a goal might be inferred to be a precondition to some subsequent action, and helpful goal adoption would transfer the speaker's goal for attaining mutual belief to the addressee. But, in general, an addressee might be able to infer only that the speaker was going to do *something* based on the outcome of a requested action, but not what that action was. In the plan-recognition models under discussion, such vague plans are not representable. Still, in the task-oriented telephone dialogue case, one would expect the apprentice to confirm success (or report trouble) even in cases where no specific plan can be attributed to the speaker.

The most promising possible explanation is to use "expectations" about the conversant' goals. In other words, there would be expectations that the apprentice would already have the goals of attaining mutual belief of the result of the requested actions. This approach may indeed work, if the notion of expectation is handled properly, but it begs the question — where would such expectations come from? This is in fact the question we are trying to answer.

Regarding the analysis of expectations *per* se, if one were to describe expectations in the Allen and Perrault model in terms of mental states, rather than in terms of data structures, they would be mutual beliefs that the conversants have a given set of goals. In the case of these task-oriented dialogues, it would be mutually believed that the apprentice has the standing goal of making public the outcome of any requested actions. But the proper mental state characterization of such expected goals cannot be based solely on mutual beliefs because those beliefs could be revised: if the apprentice did not do what was expected, the expert could simply assume his own beliefs, and hence the mutual beliefs, were wrong. Hence, the expert's expectation would disappear. Without further stipulation that the expert also wants to attain this state of mutual belief, no prediction could be made that the expert would have the goals leading to a request for confirmation, which frequently occurred in such circumstances in the corpus. But, such a stipulation would be insufficient as simple goals or desires can be changed too easily.

Both parties should not merely predict the apprentice will have the goal to attain mutual belief of successful action, they should be *committed* to his having it — the apprentice is *supposed* to confirm, and the expert can hold him to it.[3] These requirements on the apprentice arise, we argue, because both parties *jointly intend* to engage in the task. Thus, we are claiming that to characterize the nature of many situation-specific expectations properly, one needs an account of joint intention.

In an earlier papers [6, 10], we defined and explained the concepts of joint commitment and intention parallel to our treatment of individual commitment and intention [5]. In those papers, we showed how the adoption of joint commitments and intentions by agents entails their having individual commitments and intentions to do their parts of the collective activity. Here, the theory of joint intention is only briefly summarized. Then, a model of the subjects' task as a joint activity is provided, and the theory is applied to explain the origin of the goals underlying the pervasive use of confirmations. Finally, we discuss extending the theory to handle dialogue more generally as a joint activity. We now proceed to describe formally what is meant here by joint commitments and joint intentions.

2 The Formalism

Our account of individual and joint commitment and intention is formulated in a modal language of belief, goal, action, and time. Due to space limitations, we can only sketch some of the features of the formalism, and only the assumptions and general properties that are needed in the linguistic application.

In addition to the usual connectives of a first-order language with equality, we have formulas (BEL x p) and (GOAL x p) to say that x has p as a belief and goal respectively, and (MB x y p) to say that x and y mutually believe that p holds; (KNOW x p) and (MK x y p) are used for knowledge and mutual knowledge, respectively. To talk about actions, we use (DONE $x_1 \ldots x_n$ a), (DOING $x_i \ldots x_n$ a), and (DOES $x_1 \ldots x_n$ a) to say that a sequence of events describable by an action expression a was just done by the agents x_i, is being done now, or will be done 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 condition; and finally, a* is repetition. The usual programming constructs like IF/THEN actions and WHILE loops can easily be formed from these.[4] To deal with time, we use (EARLIER p),

[3] A similar situation arises in the train station domain [1]. Patrons do not simply believe that the clerk will answer questions about trains, they know he is supposed to do so.

[4] Test actions occur frequently in our analysis, yet are potentially confusing. The expression p?;a should be read as

(EVENTUALLY p), (NEVER p) and (UNTIL q p) to say that p was true at some point in the past, will be true at some point in the future, will not be true at any point in the future, and will remain true until q is true, respectively. Many of these operators can be defined in terms of the others, but that need not concern us here. For a full semantics of this language, and a discussion of its properties, see other papers of ours [5, 10].

2.1 The assumption of memory

The formalism we are developing embodies various assumptions (understood as constraints on models) concerning beliefs and goals. For example, we assume that all agents eventually drop their achievement goals by either achieving them or by giving up (see [5]), and that goals are always compatible with what is believed. In exploring the properties of joint intentions and commitments below, we also assume that individuals and groups realize what they did not believe or mutually believe in the past. More formally,

Assumption 1 *Memory*
$\models \neg(\text{EARLIER } (\text{BEL } x\ p)) \equiv (\text{BEL } x\ \neg(\text{EARLIER } (\text{BEL } x\ p)))$

That is, agent x did not believe p iff the agent now believes that he did not believe p. A corresponding property is assumed to hold for joint memory, namely:

Assumption 2 *Joint memory*
$\models \neg(\text{EARLIER } (\text{MB } x\ y\ p)) \equiv$
$(\text{MB } x\ y\ \neg(\text{EARLIER } (\text{MB } x\ y\ p)))$

In other words, we do not allow agents (or groups of agents) to have doubts or inaccurate beliefs about their past beliefs.

2.2 Individual commitment and intention

Based on these primitives, a notion of individual commitment called PGOAL, for persistent goal, has been defined [5] that describes an agent as being committed to p if he knows that he will keep his goal to eventually bring about p at least until he believes it is true, is impossible, or is irrelevant. More formally,[5]

Definition 1 $(\text{PGOAL } x\ p\ q) \stackrel{\text{def}}{=}$
$(\text{BEL } x\ \neg p) \wedge (\text{GOAL } x\ (\text{EVENTUALLY } p)) \wedge$
$(\text{KNOW } x\ (\text{UNTIL } [(\text{BEL } x\ p) \vee (\text{BEL } x\ (\text{NEVER } p)) \vee$
$(\text{BEL } x\ \neg q)]$
$(\text{GOAL } x\ (\text{EVENTUALLY } p))))$

The important points to observe about individual commitments are these: once adopted, an agent cannot drop them freely; other commitments need to be consistent with them; and agents will try again to achieve them should initial attempts fail. Condition q is an escape clause (which we will occasionally omit for brevity), against which the agent has relativized his persistent goal. Should the agent come to believe it is false, the commitment is no longer relevant, and can be dropped. Note that q could in principle be quite vague, allowing for disjunctions, quantifiers, and the like, although for sufficiently broad conditions, not much of a commitment would remain.

Merely having a commitment to get some action done is not sufficient for an agent to act deliberately. It is consistent with what we have said so far that an agent with a PGOAL to achieve (DONE x a) could blunder about at random until he discovers that he has in fact done the required action, perhaps by accident. But, an agent who does an action *intentionally* should at least realize what he is doing throughout the execution of the action. In this paper, however, we merely require the agent to believe at the *start* of the action, that he is about to do it:

Definition 2 $(\text{INTEND } x\ a\ q) \stackrel{\text{def}}{=}$
$(\text{PGOAL } x\ (\text{DONE } x\ [\text{BEL } x\ (\text{DOES } x\ a)]?;a)\ q)$

Therefore, an intention is a commitment to having done an action starting in a specific mental state.

It is also useful to model agents who have intentions where only certain parts of the overall action are specified. For example, an agent might intend to something of the form a;. . . ;b, where what is to be done between a and b is not known at the outset of the action. To express this, we use a new form of intention, INTEND*, that takes an open action expression as an argument, with the unspecified parts bound by an existential quantifier.

Definition 3 $(\text{INTEND}^* x\ \exists e.a\ q) \stackrel{\text{def}}{=}$
$(\text{PGOAL } x\ \exists e\ (\text{DONE } x\ [\text{BEL } x\ \exists e^*\ (\text{DOES } x\ a^*)]?;a)\ q)$
where e *occurs within* a, e^* *does not occur in* a, *and* a^* *is* a *with* e *replaced by* e^*.

So, for example, if we have

$(\text{INTEND}^* x\ \exists e\ a;e;b),$

then we have a commitment to there being an e such that a;e;b gets done. However, prior to executing this sequence, we do *not* require the agent to satisfy

$(\text{BEL } x\ (\text{DOES } a;e;b))$

for a specific event e, as we would with INTEND, but only that

$(\text{BEL } x\ \exists e^*(\text{DOES } a;e^*;b))$

must be true. Although the agent must know that he will do something between the a and the b, he need not know initially what it is.

2.3 Joint commitments and intentions

We have argued elsewhere [6, 10] that to act together as a team, a group of agents is in a complex mental state termed a *joint intention,* which is defined as a joint commitment to act in a shared belief state. A joint intention binds team members together, enabling the team to overcome misunderstandings and surmount obstacles. The analyses of joint commitments and intentions given in the earlier papers are motivated by nonlinguistic examples, such as driving in a convoy, and by a principle of making minimal changes to the analysis of individual intentions and commitments.

As in the individual case, we start with the notion of a joint persistent goal, JPG, which is the analogue of

"action a with p holding initially," and analogously for a;p?.

This definition differs slightly from that presented in our earlier work [5], but the difference is of no consequence here.

PGOAL with belief replaced by mutual belief, and goal replaced by MG and WMG, as below:

Definition 4

$$(\text{MG } x\ y\ p) \stackrel{\text{def}}{=} (\text{MB } x\ y\ (\text{GOAL } x\ p) \wedge (\text{GOAL } y\ p))$$

$$(\text{JPG } x\ y\ p\ q) \stackrel{\text{def}}{=}$$
$$(\text{MB } x\ y\ \neg p) \wedge (\text{MG } x\ y\ (\text{EVENTUALLY } p)) \wedge$$
$$(\text{MK } x\ y\ (\text{UNTIL } [(\text{MB } x\ y\ p) \vee (\text{MB } x\ y\ (\text{NEVER } p)) \vee (\text{MB } x\ y\ \neg q)]\ (\text{WMG } x\ y\ p))))$$

If the last line here had been (MG x y (EVENTUALLY p)), then the analogy between the individual and joint case would have been clearest. Unfortunately, we must instead use (WMG x y p) (defined below) which says that the agents have a "weak mutual goal" to achieve p. This is defined to be a mutual belief that each agent has "weak goal" to achieve p relative to the other agent, which in turn is defined as the agent either having the goal to achieve p, or, if he comes to believe (typically privately) that p is true, impossible, or irrelevant, the goal of making this a mutual belief. More precisely,

Definition 5

$$(\text{WG } x\ y\ p) \stackrel{\text{def}}{=}$$
$$[\neg(\text{BEL } x\ p) \wedge (\text{GOAL } x\ (\text{EVENTUALLY } p))] \vee$$
$$[(\text{BEL } x\ p) \wedge (\text{GOAL } x\ (\text{EVENTUALLY } (\text{MB } x\ y\ p)))] \vee$$
$$[(\text{BEL } x\ (\text{NEVER } p)) \wedge (\text{GOAL } x\ (\text{EVENTUALLY } (\text{MB } x\ y\ (\text{NEVER } p))))] \vee$$
$$[(\text{BEL } x\ \neg q) \wedge (\text{GOAL } x\ (\text{EVENTUALLY } (\text{MB } x\ y\ \neg q)))]$$

$$(\text{WMG } x\ y\ p\ q) \stackrel{\text{def}}{=} (\text{MB } x\ y\ (\text{WG } x\ y\ p\ q) \wedge (\text{WG } y\ x\ p\ q)).$$

So a joint persistent goal to achieve p relative to q means that the agents mutually believe that p is false, they mutually believe that each wants it to be true at some point, and they mutually know that they will keep p as a weak mutual goal at least until they mutually believe it holds, is impossible, or irrelevant.

This weaker notion of goal is necessary here because an agent may not be aware of what his partner has discovered privately about p; thus, it would be unreasonable to expect an agent to assume obliviously that the other is still trying to achieve p. However, the persistence of a weak goal still predicts a level of robustness: the individuals are committed to achieving p, and if they discover privately that it is done, impossible or irrelevant, they have the goal of making this mutually known. In fact, it can be shown that in normal circumstances, this goal will be a PGOAL:

Theorem 1 *(taken from [10])*

$$\models (\text{JPG } x\ y\ p) \wedge C \supset (\text{UNTIL } [(\text{MB } x\ y\ p) \vee (\text{MB } x\ y\ (\text{NEVER } p))]\ [(\text{BEL } x\ p \wedge \neg(\text{MB } x\ y\ p)) \supset (\text{PGOAL } x\ (\text{MB } x\ y\ p))])$$

So, if x and y are jointly committed to p, and some condition C holds (C says that once the agent comes to believe p, he will not change his mind), then until the agents mutually believe that p is satisfied or impossible, if one agent, say x, comes to believe privately that p holds, then he has a *persistent* goal to make p mutually believed. Similar theorems can be proven about commitments to attain mutual belief of the impossibility or irrelevance of the agreement. Thus, a JPG to achieve some condition will normally lead to a private commitment to make the outcome of that condition mutually believed.

We conclude the review of the formalism with a definition of *joint intention* that parallels exactly the definition from the individual case, replacing PGOAL by JPG and BEL by MB:

Definition 6 $(\text{JI}^* \ x\ \exists e.a\ q) \stackrel{\text{def}}{=}$
$(\text{JPG } x\ y\ \exists e\ (\text{DONE } x\ y\ [\text{MB } x\ y\ \exists e^*\ (\text{DOES } x\ y\ a^*)]?;a)\ q)$
where e *occurs within* a, e^* *does not occur in* a, *and* a^* *is a with* e *replaced by* e^*.

Thus, two agents jointly intend to do some (possibly underspecified) action iff they are jointly committed to having done the action mutually believing they were about to do it. For further discussion of how joint commitments and intentions work to bind teams together and protect them against misunderstandings, see [6, 10].

3 Commitments to Action Sequences

As we will see in the next section, our analysis of task-oriented dialogue begins by assuming that the agents jointly intend to perform together some partially specified sequence of actions. Much of our analysis depends on how a commitment to an action sequence gives rise to a commitment to elements of that sequence.

First of all, observe that if an agent is committed to doing some sequence a;b, it does *not* follow that the agent is committed to either doing a or doing b by itself. For one thing, the agent may believe that b has already been done (without being preceded by a). Also, he may only be interested in having b done just after a. Similar considerations apply to a. However, in the case of the tail of a sequence, we do get a commitment that is *relativized* to the larger goal:

Theorem 2 *Commitment to the tail of a sequence:*

$$\models (\text{PGOAL } x\ (\text{DONE } x\ a;b)) \wedge (\text{BEL } x\ \neg(\text{DONE } x\ b)) \supset (\text{PGOAL } x\ (\text{DONE } x\ b)\ [\text{PGOAL } x\ (\text{DONE } x\ a;b)])$$

$$\models (\text{JPG } x\ y\ (\text{DONE } x\ y\ a;b)) \wedge (\text{MB } x\ y\ \neg(\text{DONE } x\ y\ b)) \supset (\text{JPG } x\ y\ (\text{DONE } x\ y\ b)\ [\text{JPG } x\ y\ (\text{DONE } x\ y\ a;b)])$$

The proof is as follows: any state where a sequence a;b has just been done will necessarily be one where b was just done. Thus, a goal to achieve the former implicitly includes a goal to achieve the latter. It follows that a goal to achieve the latter must persist at least as long as a goal to achieve the former, and so the relativized version of the PGOAL (or JPG) holds.

Similar reasoning does *not* apply in general to the first element a of a sequence, even assuming that the agent believes he has never done a. The reason is that the agent may not be able to tell where a ends and b begins, but still expect to correctly execute the entire sequence. That is, an agent can start with the goal of doing a;b, then at some point, without necessarily knowing that a is done, drop the goal of doing a, but continue the sequence nonetheless, thinking that eventually a;b will have been done. For example, an agent can click on a phone receiver a number of times and know that one of those clicks disconnects the line and produces a dial tone

without ever having to know which click did it. Because the goal of doing a can be given up without thinking that it has just been done (or impossible or irrelevant), the agent is not committed to doing the action by itself.

However, if the first element of a sequence is an action that cannot be done without the agent (or agents) realizing it, then we do get an appropriate relativized commitment. In particular, this is true when the initial element of sequence is a condition requiring the agent (or agents) to believe (or mutually believe) something:

Theorem 3 *Commitment to initial belief conditions:*

$$\models (\text{PGOAL } x\ (\text{DONE } x\ [\text{BEL } x\ p]?;a) \wedge \neg(\text{EARLIER } (\text{BEL } x\ p)) \supset (\text{PGOAL } x\ (\text{KNOW } x\ p)\ [\text{PGOAL } x\ (\text{DONE } x\ [\text{BEL } x\ p]?;a)])$$

$$\models (\text{JPG } x\ y\ (\text{DONE } x\ y\ [\text{MB } x\ y\ p]?;a)) \wedge \neg(\text{EARLIER } (\text{MB } x\ y\ p)) \supset (\text{JPG } x\ y\ (\text{MK } x\ y\ p)\ [\text{JPG } x\ y\ (\text{DONE } x\ y\ [\text{MB } x\ y\ p]?;a)])$$

The proof (in the individual case) is as follows: let q stand for (BEL x p). Any state where q?;a was just done will be one where q was true earlier. So if there is a goal of having q?;a done in the future, and at that point it is believed that q has not been true, there must be a goal of q being true in the future. Now to see that a goal to achieve q must persist relative to the larger goal, imagine that at all points up to some point in the future, q?;a has remained an achievement goal, and that at no point was q thought to be true. At that point, by the memory assumption, it will be believed that q has not been true, and so there will be a goal of q being true in the future. Thus the goal to achieve q will persist as long as the one to achieve q?;a or until it is thought to be satisfied. The proof in the joint case is analogous, using the joint memory assumption.

We now proceed to show how this analysis predicts the discourse goals underlying various linguistic phenomena found in our study.

4 Modeling the Task

We have argued informally elsewhere [12] that in our telephone and audiotape conditions, the expert and apprentice jointly intended to perform the task. Moreover, it was given to the partners that the apprentice would build the pump, following part-by-part instruction from the expert. So the task for both consists of the apprentice's picking up and assembling each part in the order required by the expert. Thus, each pick-up and assembly event must occur in a context where that event is what the expert wants the apprentice to do *just then.* So if we let part be a variable ranging over parts to be assembled and ae be a variable ranging over assembly actions, which take part as an argument, then (TASK part ae) will he the full action required for that part:[6]

Definition 7 $(\text{TASK part ae}) \stackrel{\text{def}}{=}$

(GOAL exp (DOES appr (Pick-up appr part)))?;
(Pick-up appr part);
(GOAL exp (DOES appr ae))?;
ae;
(Assembled part)?

That is, the apprentice is to pick up a given part, in the context of the expert's wanting him to pick it up then. Next, he is to act on it, in the context of the expert's wanting him to do that action on that part then, after which the part will be assembled.

We can model the joint mental state that resulted from the subjects' having agreed to participate in the study, as a joint intention by the expert and apprentice to perform this task for every part:

$$\forall \text{part} \in \{\text{Parts}\},\ (\text{JI* exp appr } \exists \text{ae}.(\text{TASK part ae}))$$

Call this formula Φ. So Φ stipulates that for each part, the expert and apprentice are jointly committed to the apprentice picking up and assembling that part, and moreover, they are jointly committed to the part that is picked up being the one the expert wants just then, and the assembly action being the one the expert has selected. So although we do not stipulate the order in which the parts must be tackled, the task does *not* consist solely in somehow correctly assembling the entire pump independently of the expert.

5 Predicting the Discourse Goals Underlying Confirmations

The data show that when the conversation is proceeding smoothly, each discourse assembly segment typically has the following structure [12]: first, the expert utters a temporal marker, followed by a request for the apprentice to identify some part; the apprentice typically confirms that the identification is made, and the expert proceeds to request an assembly action to be performed; the apprentice performs the action, and then confirms that the requested action was finished. Although our account of the task in terms of Φ cannot predict what utterance events will actually take place, we can predict the presence of a number of *goals* that naturally give rise to the utterance events. We examine two of these below: confirmation of successful action and confirmation of understanding. An expanded paper will show how other patterns of dialogue follow from the analysis of joint intentions, especially requests from the expert for confirmation of referential understanding.

5.1 Confirmation of Successful Action

Given Φ, we have for each part a joint intention to execute some pick-up and assembly actions after which the part will be assembled. Expanding the definition of JI*, we have a joint commitment of the form

(JPG exp appr
∃ae.[DONE . . . ;(Assembled part)?]).

By Theorem 2 (and elimination of the quantifier), this implies that we have a commitment to getting each part assembled, relative to the larger joint commitment. That is, we have

---

We will not concern ourselves with stating that ae must be a sequence of assembly events performed by the apprentice. We will also assume that (Pick-up appr part) refers to the unique sequence of events that constitutes picking up the part by the apprentice.

(JPG exp appr (Assembled part) Φ)

Once the apprentice convinces himself that he has assembled the part correctly, this JPG dissipates since it is no longer the case that both parties mutually believe they are still trying to get the part assembled. However, by Theorem 1, the apprentice is left with a residual commitment:

(PGOAL appr (MB exp appr (Assembled part))).

It is *this* persistent goal that compels the apprentice to confirm success. Moreover, this discourse goal is a direct result of fact that the expert and the apprentice took on the assembly task as a *joint* commitment. Without that assumption, the role of the apprentice would have ended with the assembly of the part.

5.2 Confirmation of Referential Understanding

By treating the task to be performed as a joint activity based on our notion of a joint intention, we can also see where the confirmations for referential understanding originate. Again expanding the definition of JI*, we have for each part a joint commitment of the form

(JPG exp appr
  ∃ae.[DONE expr appr (MB . . .)?;(TASK part ae)]).

Now by Theorem 3 (and ignoring the unused ae quantifier), we have

(JPG exp appr
  (MK exp appr
    ∃ae* [DOES appr (TASK part ae*)]) Φ).

Because TASK is of the form $\alpha?;a$, and (DOES $\alpha?;a$) entails $\alpha \wedge$ (DOES a), an expansion of TASK in the previous joint commitment leads us to

(JPG exp appr
  (MK exp appr
    ∃ae* [(GOAL exp
      [DOES appr (Pick-up part)]) ∧ . . .]))

Concerning ourselves with the first conjunct, we can again ignore the quantifier because it plays no role. Hence, we find that under normal circumstances (namely when the first conjunct is not already mutually known), this joint commitment implies

(JPG exp appr
  (MK exp appr
    (GOAL exp [DOES appr (Pick-up part)])) Φ).

So for each part to be assembled, the conversants are jointly committed to arriving at a state where it is mutually known that the expert wants the apprentice to pick up that part at that time. In fact, at the start of the task, there are a set of such joint commitments, which will then get discharged at different times during the task.

Notice that although the part variable is quantified into this mutual knowledge, the assembly event ae never is, which means that the apprentice (in particular) does not need to know initially what assembly action is required next. This is as it should be and is a direct result of our use of JI*, which replaces the ae variable by a new one ae*, within the scope of the MK operator.

Thus, the interaction of joint intention and joint memory enables us to conclude that mutual knowledge of what actions the expert wants will occur as the dialogue progresses. This mutual knowledge is usually achieved in the corpus by the apprentice's signaling understanding with "uh-huh" or "OK,'' (where it is obvious that he has yet to do the requested action). Notice that if a simple JPG rather than a JI* were used here, it would be possible for the apprentice to simply convince the expert *after* getting the part that he indeed knew prior to the assembly what part the expert wanted him to work on.

On the other hand, if the apprentice does not yet understand what part the expert wants him to work on next, he shares with the expert a joint commitment to acquiring that knowledge, which often results in his asking *clarification questions.* For his part, the expert may attempt to attain this mutual belief with a *request confirmation of understanding,* often by way of rising intonation over noun phrases, or by explicit questions (e.g., "You see that?"). Requests for both types of confirmations can be seen in the earlier fragment.

6 Concluding Remarks: Dialogue as a Joint Activity

Many writers have argued that dialogue itself should be regarded as a joint activity (see, for example, [3, 8, 13]). What remain to be demonstrated, though, are the consequences that follow from taking this approach seriously. Which phenomena require a precise notion of collaboration for their explanation? How do collaborationist accounts of discourse predict phenomena that other theories do not? To begin to answer these questions, we have shown here how a a formal theory of joint action explains confirmations that arise in task-oriented telephone dialogues. However, we are not the only ones to have considered confirmations as evidence of joint action.

For example, regarding confirmations of understanding, SchegloffT [13] has claimed that "uh-huh"' and like utterances require treating dialogue as an "interactional achievement," an accomplishment of both conversants acting together. In his analyses, the purpose of such confirmations is to convey understanding and to signal passing up the opportunity to seek repairs or clarifications. But, such analyses are not related to other forms of confirmations, such as those of the success (or failure) of requested actions. If we are correct, in our analysis, both stem from the same underlying principles.

Clark and Wilkes-Gibbs [3] have provided an extensive set of examples of referential phenomena that, they argue, call for a collaborative explanation. Among these phenomena are sentence completions and confirmations. They claim that the key to collaborative reference is for the conversants to attain, roughly by the beginning of the next turn, a state of mutual belief that the description is adequate for present purposes. We agree with their conclusions, but observe that in their theories, conversants' goals for attaining mutual belief are not derived from more general nonlinguistic behavior. This misses

the possibility for significant generalizations.

The results obtained in this paper stem from the joint nature of the task under discussion by the conversants, and not from the joint nature of the process of engaging in a dialogue. A substantive promisory note of our approach is thus to view dialogue itself as a joint activity, one that is appropriately initiated, monitored, and closed, and is robust against miscommunication. From our perspective, by agreeing to engage in a dialogue, the conversants have tacitly adopted a joint commitment to understand one another. It is our goal to be able to model this intuition formally and derive discourse goals that underlie such phenomena as confirmations, clarifications, repairs, elaborations, and the like, for dialogue situations more general than those about joint tasks.

However, there are two reasons why we have considered only task-oriented dialogues here. First, though these dialogues are very simple in structure, there has been no satisfying account of them. We can gain much by sharpening our tools on the simple cases first.

Second, a technical difficulty is looming. We were able to derive numerous predictions about goals underlying discourse phenomena from the single assumption that the two conversants had joint commitments for the apprentice to pick up and assemble a part once he knew what part the expert wanted him work on next. Formally, this involved only quantifying over parts and events. In a more general setting, understanding what a speaker means goes beyond identifying a part or an (went, and can involve any proposition. To say that both participants are committed to making what the speaker means mutually known, we would need to quantify over the propositions. This we cannot do with our possible-worlds approach, but we look forward to a more fine-grained semantic theory (e.g., situation theory [2]) to provide the technical apparatus. Still, the present paper depicts the shape of such an account, and illustrates some of the potential benefits to accrue from treating dialogue as a joint activity.

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 [3], Grosz and Sidner [8], 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 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

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

[2] J. Barwise and J. Perry. *Situations and Attitudes.* MIT Press, Cambridge, Massachusetts, 1983.

[3] 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.

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

[5] P. R. Cohen and H. J. Levesque. Intention is choice with commitment. *Artificial Intelligence,* 42(3), 1990.

[6] P. R. Cohen and II. .J. Levesque. Teamwork. *Nous,* 25, to appear, 1991.

[7] P. R. Cohen and C. R. Perrault. Elements of a plan-based theory of speech acts. *Coqnitive Science,* 3(3):177-212, 1979.

[8] 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.

[9] J. Hobbs and D. Evans. Conversation as planned behavior. *Cognitive Science,* 4(4):349-377, 1980

[10] H. J. Levesque, P. R. Cohen, and J. Nunes. On acting together. In *Proceedings of AAA1-90,* San Mateo, California, July 1990. Morgan Kaufmann Publishers, Inc.

[11] D. J. Litrnan 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.

[12] S. L. Oviatt and P. R. Cohen. Discourse structure and performance efficiency in interactive and noninteractive spoken modalities. *Computer Speech and Language,* 1991, in press.

[13] E. A. Schegloff. Discourse as an interactional achievement: Some uses of unh-huh and other things that come between sentences. In D. Tannen, editor, *Analyzing discourse: Text and talk.* Georgetown University Roundtable on Languages and Linguistics, Georgetown University Press, Washington, D.C., 1981.