Toward a semantics for an agent communications language based on speech-acts*

Ira A. Smith and **Philip R. Cohen**
Center for Human-Computer Communication
Department of Computer Science and Engineering
Oregon Graduate Institute

Abstract

Systems based on distributed agent architectures require an agent communications language having a clearly defined semantics. This paper demonstrates that a semantics for an agent communications language can be founded on the premise that agents are building, maintaining, and disbanding teams through their performance of communicative acts. This view requires that definitions of basic communicative acts, such as requesting, be recast in terms of the formation of a joint intention — a mental state that has been suggested underlies team behavior. To illustrate these points, a semantics is developed for a number of communication actions that can form and dissolve teams. It is then demonstrated how much of the structure of popular finite-state dialogue models, such as Winograd and Flores' *basic conversation for action*, follows as a consequence of the logical relationships that are created by the redefined communicative actions.

Introduction

A language for interagent communication should allow agents to enlist the support of others to achieve goals, to commit to the performance of actions for another agent, to monitor their execution, to report progress, success, and failure, to refuse task allocations, to acknowledge receipt of messages, etc. Crucially, a collection of agents needed to accomplish a task will frequently include humans who have delegated tasks to the agents, and/or humans who will be performing some of the work. As such, it is essential that the *functions* being offered by the communication language be common across the language of intelligent agents and the language that people will use to communicate with them.[1] It so happens that there is such a language, the language of "speech acts" (Austin 1962; Searle 1969), or more precisely, "illocutionary acts." Such actions include requesting, promising, offering, acknowledging, proposing, accepting, etc.

Recently, a number of researchers have proposed artificial languages based on speech act theory as the foundation for interagent communication (Labrou & Finin 1994; External Interfaces Working Group 1993; Shoham 1993; Sidner 1994). The most elaborate and developed of these is KQML (External Interfaces Working Group 1993). In this language, agents communicate by passing so-called "performatives" to each other. KQML is offered to the agent community as an *extensible* language with an open-ended set of performatives, whose meaning is independent of the propositional content language (e.g., Prolog, first-order logic, SQL, etc.) However, KQML does not yet have a precise semantics.[2] Without one, agent designers cannot be certain that the interpretation they are giving to a "performative" is in fact the same as the one some other designer intended it to have. Moreover, designers are left unconstrained and unguided in attempts to extend the set of communication actions, and they cannot be certain how the existing or newly defined communication acts will be used in interagent communication protocols.

This paper takes a first step to rectify this situation by illustrating how communicative actions can be given semantics and how this semantics can be used to analyze interagent communication protocols. To provide a basis for the development of such protocols, we propose that the purpose of agent communication is to achieve tasks through the creation, monitoring, and disbanding of teams. Teamwork is founded on the notion of joint intentions (Cohen & Levesque 1991; Levesque, Cohen, & Nunes 1990). Thus, the purpose of these communication actions we provide will be to create, monitor, and discharge joint intentions among the agents involved.

The organization of the paper is as follows: first we review a formal theory of teamwork and of speech acts. Then we present new definitions of communicative ac-

This is supported by the Information Technology Promotion Agency, Japan, as a part of the Industrial Science and Technology Frontier Program " New Models for Software Architectures " sponsored by NEDO (New Energy and Industrial Technology Development Organization) and by the Advanced Research Projects Agency (contract number DABT63-95-C-0007) of the Department of Defense. The results presented here do not reflect the position or policy of the either the Japanese Government or the US Government.

[1] Note that we say "function", rather than "expression" in that obviously, people do not speak artificial languages.

[2] A first attempt has been made (Labrou & Finin 1994), but much work remains.

tions sufficient to form, maintain and disband teams. Finally, the framework is shown to provide a basis for analysing finite-state dialogue models.

Modeling Teams

We regard team activity as the key to agent interactions. Team activity is more than having mutual goals, coordinating actions, or even being mutually helpful (Grosz & Kraus 1993; Grosz & Sidner 1990; Searle 1990; Tuomela & Miller 1988). A team must be able to hold together in the face of adverse conditions, individual doubt, and communication failure. Team members commit resources to form teams, and need to recover those resources when the team disbands. Following (Cohen & Levesque 1991), we shall say that a team is formed when the agents have a joint commitment or joint intention with the other agents forming the team with respect to a specific goal. This concept of a joint intention, the heart of a team, is defined in terms of the individual team members commitments and weak achievement goals. Below is a brief overview of a model of agents upon which a model of joint action is built.

The formal framework for modeling team behavior given below should be regarded as a specification, rather than as a set of formulas that agents themselves are supposed to manipulate. Various implementations of cooperating agents have been inspired by the specification (Jennings & Mamdani 1992; Tambe 1996)

Syntax

The language we 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 p follows from x's beliefs or goals (a.k.a choices) respectively; (AGT x e) says that x is the only agent for the sequence of events e; $e_1 \leq e_2$ says that e_1 is an initial subsequence of e_2; and finally, (HAPPENS a) and (DONE a) say that a sequence of events describable by an action expression a will happen next or has just happened, respectively. Versions of HAPPENS and DONE specifying the agent (x) are also defined. AFTER, LATER, BEFORE, and PRIOR are defined using HAPPENS. Knowledge (KNOW) and the various types of mutual belief (ABEL,MB, and BMB) are defined in a standard manner. Details of the semantics can be found in (Cohen & Levesque 1990a).

An action expression here is built from variables ranging over sequences of events using the constructs of dynamic logic (Harel 1979): 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 such as IF/THEN actions and WHILE loops, can easily be formed from these. Because test actions occur frequently in our analysis, yet create considerable confusion, read p?;a as "action a occurring when p holds," and for a;p?, read "action a occurs after which p holds." We use e as a variable ranging over sequences of events, and a and b for action expressions.

Individual Commitments and Intentions

We define an agent's commitment to be a goal that is kept sufficiently long — a persistent goal. Intention is a kind of persistent goal in which an agent commits to having done an action believing he is about to do it. Again the reader is referred to (Cohen & Levesque 1990b) for the full semantics. A joint persistent goal is used to bind agents through mutual commitment.

Definition 1 *Joint Persistent Goal*

$$(\text{JPG x y p q}) \stackrel{\text{def}}{=} (\text{MB x y } \neg\text{p}) \wedge (\text{MG x y p}) \wedge (\text{BEFORE } [(\text{MB x y p}) \vee (\text{MB x y } \Box\neg\text{p}) \vee (\text{MB x y } \neg\text{q})] \ (\text{WMG x y p}))$$

where:
Mutual Goal is defined to be

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

Weak Achievement Goal is:

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

Weak Mutual Goal is:

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

Agents are each committed to the goal p. Moreover, if p involves the agents' each performing individual actions, then each agent is committed to the other's success. The characteristics embedded in the WAG and JPG are necessary to allow teams to balance the responsibilities an individual has towards the team with the requirement that an individual team member be allowed to drop a goal under certain reasonable conditions (Cohen & Levesque 1991). If an agent discovers the goal has been accomplished or is impossible, or if he discovers the relativizing condition is no longer true, he is allowed to drop the goal. However, the agent is still left with a goal to make his discovery mutually believed by the rest of the team. The details of these definitions can be found in (Levesque, Cohen, & Nunes 1990; Cohen & Levesque 1991).

To these definitions we add:

Definition 2 *Persistent Weak Achievement Goal*

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

The PWAG requires more commitment from an agent then is required by a WAG. Upon discovering that p has been achieved or has become impossible, or that q is no longer true, the agent will be left with a PGOAL to reach mutual belief with the other team members about the status of p or q. We will use the PWAG during team formation. The following proposition follows directly from the definitions of WAG and PWAG.

Proposition 1

(PWAG x y a p) ⇒ (WAG x y a p)
Moreover, we have shown in (Cohen & Levesque 1991) that:

Proposition 2

(JPG x y p q) ⇒ (PWAG x y p q)

Communicative Acts

A language of communicative acts will serve as the means that agents use to communicate their mental states and to form teams with other agents to achieve their goals. Other agent communication languages based on communicative acts exist, probably the best known example of which is KQML (Finin *et al.* 1994; Labrou & Finin 1994). It was argued in (Cohen & Levesque 1995) that despite the long list of so-called "performatives", KQML contained essentially two types of communicative actions — requestives and assertives, with a wide variety of types of contents. For the present, we also employ these two act types.

Attempts

By performing a particular communicative act, an agent is deliberately *attempting* to alter the state of the world. In doing so there is a specific result the agent desires the act to accomplish that is related to one or more of the agent's goals. However, because the communicating agent's *goal* is to alter the receiving agent's mental state in a particular way, there is no guarantee the act will achieve this result. For example, if I ask you to close the door, I may be sure that you heard and understood me, but there is no guarantee you will close the door. Therefore, to characterize an attempt we specify, in addition to the agent's actual goal, a minimum achievement to which the agent is committed. More specifically, an attempt requires four arguments: the agent, the act to be performed, the goal of the attempt, and the minimal result to which the agent is committed. The following formalization of an attempt is from (Cohen & Levesque 1990a).

Definition 3 *Attempt*

{ATT x e q p} $\stackrel{def}{=}$
  [(BEL x ¬p ∧ ¬q) ∧
    (GOAL x e;q?) ∧ (INT x e;p?)]?; e

Event e is an attempt to achieve q with minimal acceptable results p iff the agent believes q not to be true at the present time and wants e to bring about q. Whereas the agent may only have a limited commitment to q, she has the intention of achieving at least p. If she were to come to the conclusion that the attempt failed to achieve even this, we could predict the agent would reattempt; that is she would either perform e again or perform a similar action to achieve the same result. All communicative actions will be defined herein as attempts.

Request

An agent will use the request speech act to attempt to induce another agent to perform a task. Often this will be a task that fits as a subtask in the overall plan of the requesting agent.

Definition 4 *Request*

(REQ x y e a p) $\stackrel{def}{=}$(ATT x e ϕ ψ)
where ϕ is:
◊(DONE y a) ∧
  (PWAG y x (DONE y a) [PWAG x y (DONE y a) p])
and ψ is :
(BMB y x (PWAG x y [(DONE y a) ∧
                    (PWAG y x (DONE y a)
                      (PWAG x y (DONE y a) p))]))

The goal (ϕ) of a request consists of two parts, the first is the straightforward requirement that the addressee perform the requested act. The second part of the goal places a requirement on the addressee's mental state, namely that y should not only intend to perform a, but perform it with respect to x's PWAG that y do it (relative to p). If y were to do a accidentally the act would not meet the goal of x's request because the requisite mental state would be absent.

The minimum result x is committed to is ψ — y's believing that it is mutually believed that x has a persistent weak achievement goal that both she eventually do a, and that she have a weak achievement goal to eventually do a relative to x's PWAG . Should x come to believe that even this result has not been achieved, by the definition of an attempt we would expect x to redo the request.

Unlike other definitions of requesting, (Cohen & Levesque 1990b) our definition is motivated by the formation of a team, which in turn is based on the notion of joint persistent goal. JPG is defined in terms of mutual belief in the existence of each individual team member's WAG. Having publicly committed to the PWAG the requester has informed the requestee that he has the WAG. Thus the requester has *already* made the individual commitments required for the formation of a team, he is already treating the requestee as a team member. Although this requirement forces the requesting agent to commit resources to team obligations, the agent receiving the request is as yet under no such obligation. The requesting agent is therefore expending resources to form a team, and is committed to a future expenditure of resources, however minimal, to maintain the team. This commitment is practical because the requesting agent is able to assume the addressee will notify him with either a confirmation or a refusal.

From the definition of a request, we can prove the requester has a persistent goal to achieve a. Our chain of reasoning will be based on an assumption of sincerity and on the definition of a weak achievement goal.

Assumption 1 *Sincerity*

⊨(∀x ∈ agent, ∀e ∈ events)

(GOAL x [HAPPENS x e;(BEL y p)?]) ⇒
(GOAL x [HAPPENS x e;(KNOW y p)?])

The sincerity assumption requires an agent who has a goal that another agent (y) come to believe a proposition (p), also should have the goal that y come to know p. This assumption asserts that no agent wants another agent to believe a proposition falsely. This implies that agents can be trusted and are trusted by each other. When an agent receives a message, she can assume the message was sent in good faith[3]. The assumption does not insist that agents be infallible — it is possible for an agent to be wrong in its beliefs, and to send messages that reflect those mistaken beliefs.

Proposition 3

⊢(HAPPENED (REQ x y c a p)) ⇒
(PWAG x y (DONE y a) p)

Proof Sketch: By assumption x is sincere. Since x is committed to establishing a belief by y that x has a PWAG for y to do a, x must have the PWAG that y do a.

Assert

The second basic communicative action we will need is assertion. ASSERT is used by an agent to attempt to "make public" that a particular proposition q is true by getting the addressee to believe that it is mutually believed that the speaker believes it to be true.

Definition 5 *Assert*

{ASSERT x y e p } $\stackrel{\text{def}}{=}$
{ATT x y e
[BEL y (BEFORE e (BEL x p))]
[BMB x y
(BEFORE e [GOAL x
(AFTER e
[BEL y (BEFORE e (BEL x p))])])]
}

That is, an assertion e from x to y that p is an attempt whose goal is to achieve the addressee y's believing that the speaker x believed p while the speaker is intending to make public what mental state the speaker *was* in — i.e., to make it mutually believed that before the attempt the speaker wanted that after the event the addressee would believe the speaker believed p. This definition follows Grice (Grice 1957) — the desired effect is achieved by means of the recognition of the speaker's intention to do so. In fact, we can prove the following:

Theorem 1

(MK x y (HAPPENED (ASSERT x y e p))) ⇒
(BEFORE e (BEL x p))

Proof: From the definition of assertion,
(BMB y x (HAPPENED (ATT x y e Φ Ψ)))
where Φ is:
(GOAL x e; (BEL y (BEFORE e (BEL x p))))
and Ψ is:

(INT x e;
[BMB y x
(BEFORE e
(GOAL x
(AFTER e (BEL y(BEFORE e (BEL x p)))))])

Thus, the intention Ψ is satisfied because there is now (after the event e) a BMB that before e Φ. In other words, the speaker's minimal intention has succeeded, and he has conveyed his mental state, which is that he wanted that y should come to believe that he believed p before doing e. Thus, we have:
(GOAL x (AFTER e (KNOW y
(BEFORE e (BEL x p))))).
Since knowledge entails truth, and (AFTER e (BEFORE e α)) entails α, we have that (BEFORE e (GOAL x (BEL x p))), which entails (BEL x p). Thus, we have in fact (MK x y (BEFORE e (BEL x p))).
Q.E. D.

If we make the further assumptions that the utterance e does not change the truth value of its content, and that agent's beliefs are persistent by default (choose your theory), then one can derive that (BEL x p) now.

Thus, if the network functions, communication channels are reliable, the speaker is sincere, and the speech act does not make its content false, mutually knowing that an assertion has occurred entails mutually knowing that the speaker believes it.

Refuse

Receipt of a request does not commit the receiving agent to accepting it — the agent may refuse a request for any reason, such as a prior conflicting goal. A refusal will be modeled as an assertion that the agent is not committed to the goal of a prior request (expressed as a PWAG).

Definition 6 *Refuse*

(REFUSE y x e a p) $\stackrel{\text{def}}{=}$(ASSERT y x e ϕ)
where ϕ is
□¬(PWAG y x (DONE y a) ψ)
and ψ is
(PWAG x y (DONE y a) p)

Unlike the request, where the requester is attempting to have the addressee take on a particular mental state, that of commitment to a future action that will have an associated cost, the refusal is simply an attempt by the original requestee to make known to the original requester that she will not ever commit to the requested action. Thus, a result of a refuse is the *requesting* agent is freed of the team obligations that were incurred by making the original request – the requester can drop the PGOAL that was embedded in her PWAG.

Theorem 2

⊢ (HAPPENED [(REQ x y e a p); (REFUSE y x e a p)])
⇒ ¬(PGOAL x (DONE y a) p)

[3]One can model situations in which this assumption is not in force.

Proof Sketch: Essentially, the content of y's refusal makes impossible the requestor's PGOAL that y do a.

Y's refusal tells x there is no team, and frees him from any commitments toward y with respect to the original request. However, the refusal may have no effect on x's PGOAL of a's being done, if x has such a PGOAL that is independent of his goal of (DONE y a). If this is the case, we would expect x to continue to pursue the achievement of a by some other means than that of y's cooperation.

Confirm

The confirm speech act is used by an agent to notify a requesting agent that she is accepting the weak achievement goal in the request. By accepting the requester's PWAG, the speaker is committed to do a and is also committed to the other obligations of team membership.

Definition 7 *Confirm*

(CONFIRM y x e a p) $\stackrel{\text{def}}{=}$(ASSERT y x e ϕ)
where ϕ *is*
(BMB x y (PWAG y x (DONE y a)
(PWAG x y (DONE y a) p)))

As was the case with REFUSE we can define a CONFIRM as an assertion with a specific type of propositional content.

A set of propositions that are analogous to those of the request speech act hold for confirm.

Proposition 4

$\vdash$ (DONE (CONFIRM y x e a p)) $\Rightarrow$
(PWAG y x (DONE y a) (WAG x y (DONE y a) p))

A proof for this proposition is omitted as it is analogous to that of proposition 3.

Building and Disbanding Teams

Now that we have sketched a semantics for the communicative acts, we will show how these acts are used to create and dissolve teams. Under normal circumstances, a request followed by a confirm will establish a joint persistent goal between x and y, relative to x's PWAG, to achieve a.

Building a team

Theorem 3

$\vdash$ (MK x y[HAPPENED (REQ x y e a p);
(CONFIRM y x e1 a p)]) $\Rightarrow$
[JPG x y (DONE y a) (PWAG x y (DONE y a) p))]

Proof sketch:

From the definition, to prove the JPG exists we must show three conditions are true:

A x and y must mutually believe (DONE y a) is currently false.

B x and y must believe the agents want (DONE y a) to eventually be true, and this must also be mutually believed.

C They must have a weak achievement goal for $\Diamond$(DONE y a), relative to x's request; and they must mutually know that this condition will hold for each of them until they mutually believe the goal is true or it will never be true or the REQ has been withdrawn.

Disbanding the team

Having created a team, we must supply a method to dissolve it once the JPGs goal has been accomplished. Just as team creation is a process that builds interlocking PWAGs into a JPG, dissolving the team is a process that unwinds the PWAGs and the JPG. This process is accomplished with a series of speech acts.

The ASSERT will be used is to inform an agent that a previously requested goal has been accomplished. A series of assertions by the team members will allow a team to be disbanded. If the ASSERT succeeds in achieving its goal, that is if the addressee comes to believe the goal has been achieved, the PGOAL associated with the assertion is dropped.

Theorem 4

$\vdash$ (MK x y [HAPPENED (REQ x y e0 a p);
(CONFIRM y x e1 a p);a;
(ASSERT y x e3 (DONE y a));
(BEL x (DONE y a))?])$\Rightarrow$
[$\neg$(PGOAL x y (DONE y a) p)
$\wedge$ $\neg$(PGOAL y x (DONE y a) p)]

Proof:

1. From Theorem 3 we know that after the CONFIRM there exists a JPG based on x's REQ .
2. From Proposition 1 (BEL y (DONE y a)) is true after the ASSERT .
3. Because PGOALs are achievement goals, step 2 requires x and y to drop the PGOAL of (DONE y a).

It is important to note that although the team members can drop their individual persistent goals that y do a, the necessary conditions that will allow the team to be completely disbanded have not yet been achieved. Each of the team members are left with the requirement to establish mutual belief that the goal has been accomplished. Until this occurs, the team members still have obligations from their PWAGs and their WMG in the JPG. One way to establish mutual belief will be for x to assert that she also believes y has achieved the goal.

Theorem 5

$\vdash$ (MK x y [HAPPENED (REQ x y e0 a p);
(CONFIRM y x e1 a);a;
(ASSERT y x e3 (DONE y a));
(BEL x (DONE y a))?;
(ASSERT x y e4 (BEL x (DONE y a)))]) $\Rightarrow$
(MB x y (DONE y a)) $\wedge$ $\neg$(JPG x y (DONE y a))

Proof Sketch: Y's assertion that she has done a produces (MK x y (BEL y (DONE y a))). X's assertion that he believes it produces the required mutual knowledge that discharges the joint commitment.

Application of the Theory

Based on observed regularities in human dialogue such as: questions are usually followed by answers, requests are usually followed by acceptances (here, termed confirmations) or refusals, many researchers have asserted that interagent dialogue follows a finite-state grammar model (e.g., (Winograd & Flores 1988)). Although there are concerns about the adequacy of this model for human interactions (Cohen 1994), it may be an adequate model for a communications protocol among software agents. Indeed, numerous researchers have advocated it, most recently, (Finin *et al.* 1994; Bradshaw *et al.* 1995). Even in the limited context of communications among software agents, the finite-state models have a major shortcoming in that no definitions are usually given for the transition types, no reasons are provided for the observed transitions between states of the model and no explanation is given of why or how a particular subgraph constitutes a dialogue unit. Our formalization of communicative acts based on joint-intention theory provides an initial explanation for the states, the state transitions, and their combination as finite-state dialogue models. As an illustration of this, we will examine Winograd and Flores' *basic conversation for action*, showing how the behavior of this dialogue model is explained by our model of interagent communication.

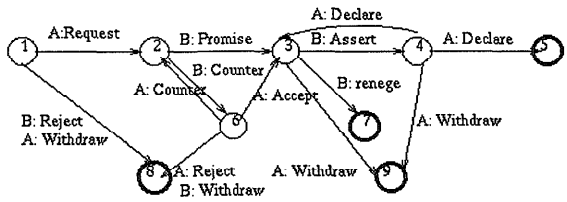

Figure 1: Winograd & Flore's basic conversation for action

In the diagram representing the conversation, (Figure 1), the nodes (circles) represent the states of the conversation, and the arcs (lines) represent speech acts that cause transitions from state to state in the conversation. Winograd and Flores assert that states 5, 7, 8 and 9 represent final states of the conversation, state 1 initiates the conversation and the other states (2, 3, 4 and 6) represent an intermediate stage of the dialogue.

The initial request, moving the dialogue from state 1 to state 2 is represented as a request of the form (REQ A B e p q) in our model. A is requesting that B perform some action p with respect to a relativizing condition q, which for us, starts a task-oriented dialogue whose purpose is to build a team to accomplish p. By performing the REQ, A has committed resources towards team formation. From proposition 3, A has an outstanding PWAG and hence a PGOAL. The model requires that the agent continue to pursue these goals until either they are accomplished or the agent believes they impossible or irrelevent. States in which A achieves one of these conditions will be the final states of this dialogue. The conversational units in this dialogue model will be those paths leading from the start state to one of the final states in the diagram.

The remainder of this section will examine the paths through the diagram that start at the initial state and end in some final state. The important paths will be analyzed with regard to the process of team formation, maintenance and disolution[4]. In addition we will be able to characterize the type of completion (successful or unsuccessful) with regard to the completion of the original task that a conversation following the path achieved.

The paths representing the hearer rejecting the task are the paths leading to state 8, namely $1-2-8$ and $1-(2-6)^*-2^*-8$. The first of these is clearly the exact sequence of acts that allow the application of Theorem 2. The result of this theorem is the requester no longer has any obligations toward potential team members. The hearer's immediate refusal means he never assumed any obligations. No explicit withdrawal speech act is needed, as claimed by W&F, as there are no outstanding persistent goals engendered by the joint commitment.

An analysis of the other paths leading to state 8 requires a definition of a counteroffer. A counteroffer is a series of events that express two speech acts. The original hearer is saying that he refuses the original request, but if the original requester would make a (slightly) different request the hearer will commit to achieving it.

Definition 8 *Counteroffer*

(CO x y a a1 q) $\stackrel{\text{def}}{=}$(REF x y e a);(ASSERT x y e1 ϕ)
where ϕ is:

$\forall$e2.[(DONE (REQ y x e2 a1 q)) $\Rightarrow$
(DONE e2;(PWAG x (DONE x a1)
(WAG y x (DONE x a1) q))?)]

The context for a counteroffer is a WAG that was communicated to the party making the counteroffer. In this dialogue, the WAG is supplied by the original request or by a previous counteroffer in the case of a counter-counteroffer. In all cases a series of counteroffers followed by a rejection allow the application of Theorem 2. The counteroffer illustrates an important point in the design of an agent communication language. There is no need to have a large set of primitive communication acts in the language. When new capabilities are required, they can be built by sequencing (e.g. COUNTEROFFER), by specialization of content (e.g. CONFIRM) or by embedding of actions from

[4]Space precludes an explicit analysis of all the paths.

the existing set of communication acts. The reader is referred to (Cohen & Levesque 1995) for additional examples of the compositionality of these communication acts.

Since all the paths that lead to state 8 leave all parties free of team obligations and all goals discharged, 8 is a final state. The dialogue has been brought to a successful conclusion. However, state 8 represents a failure in that the team was never fully formed, and the task was left unaccomplished.

All other paths to final states lead through state 3. The simplest path to 3 is $1-2-3$. This path is a REQ followed by a CONFIRM . As we have shown in theorem 3, this establishes a JPG for A and B. The other path to 3 is $1-(2-6)^*-3$. We have already analyzed the first two sets of links in this path, the arc from $6-3$ is labeled with A:Accept. A is accepting B's counteroffer, that is A is performing (REQ A B e a1 q). Note that an (ASSERT A B (WAG A B (DONE B a1) q)) will accomplish the same goals implying that either could implement this transition. Action a1 and q are bound by B's counteroffer. B's counteroffer followed by A's acceptance (i.e., the REQ or an ASSERT that A wants B to do a1) have created a set of interlocking WAGs, which establish a JPG. All paths leading to 3 are paths where a JPG exists as the state is entered. This state represents a point where Theorem 3 applies. A team has been formed, both A and B have undischarged commitments as the result of the JPG. Any path leading to a final state from state 3 must provide a way to discharge both sets of commitments.

The shortest paths out of state 3 are those leading directly to states 7 and 9. One is labeled B:Renege, the other A:Withdraw. In the case of the Renege, B is performing (ASSERT B A $\Box\neg$(GOAL B $\Diamond$(DONE B a)) (we assume it is obvious that the goal has not been achieved, but is still possible). From this we know $\Box\neg$(GOAL B $\Diamond$(DONE B a)), and since A is introspectively competent, and A does too. This forces the two parties to discharge the JPG and disband the team.[5]

Similarly, the arc labeled A:Withdraw represents (ASSERT A B $\neg$(WAG A B (DONE A a) q)). Receipt of this communication allows B to discharge his WAG, as the original JPG was created relative to A's original PWAG.

The remaining paths from state 3 require the analysis of the $(3-4)^*$ segment. In this segment, B is performing (ASSERT B A (DONE B a))—that is, B is claiming to have finished the task. The arc from state 3 to 4 is A disagreeing—(ASSERT A B $\neg$(DONE y a)). The paths $3-(4-3)^*-7$ and $3-(4-3)^*-9$ have the same meaning as the corresponding paths without the $(4-3)^*$ segment.

We have finished examining all the paths into states 7 and 9, in all cases the goals that were active in state 3 have been discharged upon entering these states. These final states represent conditions under which a team was formed, and then was later disbanded without accomplishing its task.

The last path to be analyzed is the one leading from state 3 to 5. The only segment that remains to be discussed is $4-5$. On this arc A is communicating (ASSERT A B (BEL A (DONE B a))), that is A is agreeing with B's prior assertion. This represents the situation described in theorem 5. All goals have been discharged, and the team is disbanded with its task accomplished.

In summary, the formation and disbanding of teams via communicative actions can explain such dialogue protocols. We have shown that some of the arcs in the diagram are correct, some incorrect, and others are missing. (In fact, it now makes sense to talk about arcs that are missing for a dialogue specification.) The methodology developed here could now be used to analyze the specifications given for other protocols, such as the "contract net" (Smith & Davis 1988).

Conclusions and Future Work

This paper has attempted to show that we can define communicative acts in terms of the mental states of the agents performing the act, and that these acts are effective ways to form, regulate and disband teams of agents. The mental states of the agents include the commitments the agents in a team have toward each other and toward the team's task. These communications acts and the mental states they represent can be used as the basis for an agent communication language's semantics. We have applied the theory to a model of interagent protocol, and shown the theory explains the structure of that protocol. In the process we have demonstrated that our small set of primitive acts can be composed to define more complex communicative acts.

Our policy of building new operators from an existing set of well-defined primitives leads to a consistent and well-understood semantics for the language. Furthermore, it offers the possibility that agents can themselves enlarge the set of communicative actions by decomposing non-primitive ones into their more primitive parts.

References

Austin, J. 1962. *How to Do Things with Words.* Clarendon Press.

Bradshaw, J.; Dutfield, S.; Benoit, P.; and Wooley, J. D. 1995. KAoS: Toward an industrial-strength open agent architecture. In Bradshaw, J. M., ed., *Software Agents.* AAAI/MIT Press.

Cohen, P. R., and Levesque, H. J. 1990a. Intention is choice with commitment. *Artificial Intelligence* 42:213–261.

[5] In fact, in the joint intentions analysis, this is not quite true. If an agent changes its mind, one would ideally like to say that the JPG has been "violated." The model can so far only say that there was not in fact a JPG to begin with. Fixing this is an active subject of our present research

Cohen, P. R., and Levesque, H. J. 1990b. Rational interaction as the basis for communication. In Cohen et al. (1990). chapter 12, 221–256.

Cohen, P. R., and Levesque, H. J. 1991. Teamwork. *NOÛS* 21:487–512.

Cohen, P. R., and Levesque, H. J. 1995. Communicative actions for artificial agents. In *Proceedings of the First International Conference on Multi-Agent Systems.* AAAI Press / MIT Press.

Cohen, P. R.; Morgan, J.; and Pollack, M. E., eds. 1990. *Intentions in Communication.* System Development Foundation Benchmark Series. Bradford Books, MIT Press.

Cohen, P. R. 1994. Models of dialogue. In Ishiguro, T., ed., *Cognitive Processing For Voice and Vision.* Society of Industrial and Applied Mathematics.

External Interfaces Working Group, A. K. S. I. 1993. Specification of the KQML agent-communication language. Working paper.

Finin, T.; Fritzson, R.; McKay, D.; and McEntire, R. 1994. KQML as an agent communication language. In *Proceedings of the Third International Conference on Information and Knowledge Management.* ACM Press.

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

Grosz, B., and Kraus, S. 1993. Collaborative plans for group activities. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence*, 367–373. IJCAII.

Grosz, B., and Sidner, C. 1990. Plans for discourse. In Cohen, P. R.; Morgan, J.; and Pollack, M. E., eds., *Intentions in Communication.* Cambridge, Massachusetts: MIT Press. 417–444.

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

Jennings, N. R., and Mamdani, E. H. 1992. Using joint responsibility to coordinate collaborative problem solving in dynamic environments. In *Proceedings of the Tenth National Conference on Artificial Intelligence*, 269–275. Menlo Park, California: American Association for Artificial Intelligence.

Labrou, Y., and Finin, T. 1994. A semantics approach for kqml – a general purpose communication language for software agents. In *Proceedings of the Third International Conference on Information and Knowledge Management.* ACM Press.

Levesque, H. J.; Cohen, P. R.; and Nunes, J. H. T. 1990. On acting togcther. In *Proceedings of the Eighth National Conference on Artificial Intelligence*, 94–99. AAAI.

Searle, J. 1969. *Speech Acts.* Cambridge University Press.

Searle, J. 1990. Collective intentions and actions. In Cohen et al. (1990). chapter 19, 401–415.

Shoham, Y. 1993. Agent-oriented programming. *Artificial Intelligence* 60(1):51–92.

Sidner, C. 1994. An artifical discourse language for collaborative negotiation. In *Proceedings of the National Conference on Artificial Intelligence (AAAI'94)*, 814–819. AAAI Press.

Smith, R. G., and Davis, R. 1988. Negotiation as a metaphor for distributed problerm solving. In Bond, A. H., and Gasser, L., eds., *Readings in Distributed Artificial Intelligence.* San Mateo, Caifornia: Morgan Kaufmann. 333–356.

Tambe, M. 1996. Tracking dynamic team activity. In *Proceedings of the National Conference on Artificial Intelligence.* AAAI.

Tuomela, R., and Miller, K. 1988. We-intentions. *Philosophical Studies* 53:367–389.

Winograd, T., and Flores, F. 1988. *Understanding Computers and Cognition.* Addison-Wesley.