Toward a Semantics for a Speech Act Based Agent Communications Language*

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

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

September 18, 1995

Abstract

Implementations of systems based on distributed agent architectures require an agent communications language that has a clearly defined semantics. Without such a semantics, neither agents nor developers can be sure what another agent's commitment to perform a task means. Agent to agent communications can be viewed as a task oriented dialogue for building, maintaining and disbanding teams. A speech act model based on joint intentions can provide the means for agents to build maintain and disband teams. It can describe the outward appearance of a task oriented dialogue, and can provide motivation for the step by step structure of the dialogue. In this paper we begin to develop a semantics for an agent communications language based on the interactive abilities of agents. We use this semantics to demonstrate how teams of agents can be formed and dissolved, and to analyze the Winograd & Flores *basic conversation for action.*

1 Introduction

The paradigm of distributed agent computing presumes problem solving through interacting agents. This collaboration requires a language for agent communications that must have a clear semantics to allow the agents and the agent developers to understand what an agent's communications mean. In this paper we begin to develop a semantics for an agent communications language based on the interactive abilities of agents and use this semantics to analyze finite state protocols such

*This research was supported by a grant from the New Energy Development Organization (Japan) 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.

Relativized Persistent Goal:

$$(\textbf{PGOAL}\ x\ p\ q) \triangleq (\textbf{GOAL}\ x\ p) \wedge (\textbf{BEFORE}\ [(\textbf{BEL}\ x\ p) \vee (\textbf{BEL}\ x\ \Box\neg p) \vee (\textbf{BEL}\ x\ \neg q)]\ \neg(\textbf{GOAL}\ x\ (\textbf{LATER}\ p)))$$

Mutual Goal:

$$(\textbf{MG}\ x\ y\ p) \triangleq (\textbf{MB}\ x\ y\ (\textbf{GOAL}\ x\ \Diamond p) \wedge (\textbf{GOAL}\ y\ \Diamond p)$$

Weak Achievement Goal;

$$(\textbf{WAG}\ x\ y\ p\ q) \triangleq [\neg(\textbf{BEL}\ x\ p) \wedge (\textbf{GOAL}\ x\ \Diamond p)] \vee [(\textbf{BEL}\ x\ p) \wedge (\textbf{GOAL}\ x\ \Diamond(\textbf{MB}\ x\ y\ p))] \vee [(\textbf{BEL}\ x\ \Box\neg p) \wedge (\textbf{GOAL}\ x\ \Diamond(\textbf{MB}\ x\ y\ \Box\neg p))] \vee [(\textbf{BEL}\ x\ \neg q) \wedge (\textbf{GOAL}\ x\ \Diamond(\textbf{MB}\ x\ y\ \neg q))]$$

Weak Mutual Goal:

$$(\textbf{WMG}\ x\ y\ p) \triangleq (\textbf{MB}\ x\ y\ (\textbf{WAG}\ x\ y\ p) \wedge (\textbf{WAG}\ y\ x\ p))$$

Relativized Intention:

$$(\textbf{INT}\ x\ a\ p\ q) \triangleq (\textbf{PGOAL}\ x\ (\textbf{DONE}\ x\ (\textbf{BEL}\ x\ (\textbf{HAPPENS}\ x\ a))?;a)\ p)$$

Figure 1: Definitions of selected mental states

as Winograd & Flores' *basic conversation for action*. Because the language provides operators corresponding to agent's communicative capabilities, its semantics is based on an underlying theory of agent interaction which we claim is best modelled as a joint activity [CL90b, CL90a, Coh94].

2 The Theory of Joint Intentions

This theory describes agent behavior in terms of characterizations of an agent's internal state. The reader is referred to [CL90b, CL90a, Coh94] for details of the theory. An agent's internal state is described as a modal logic with a dynamic logic of actions. Some of the less familiar mental states are summarized in Figure 1

Two key concepts are the *relativized persistent goal* (**PGOAL**)and the *relativized intention* (**INT**). An agent with a **PGOAL** of p has a goal of p and has a commitment to the goal. The agent cannot drop the goal unless she believes it accomplished or impossible, or unless she believes a relativizing condition (q) to be untrue. A relativized intention is a persistent goal to perform an action while in a particular mental state. The agent is committed to being in a state where she believes she is about to do the intended action next.

3 TEAMS

A fundamental point of joint intention theory is joint action is more than just a coordination of the actions of the individual team members. In addition to coordination of actions, there must also be a coordination of the agents' mental states. The team is formed when each agent has a 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 intentions and **WAG**s.

Definition 1 *Joint Persistent Goal*

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

The characteristics embedded in the **WAG** and **JPG** are necessary to allow teams to have coherence by balancing 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 [CL91]. 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 [LCN90, CL91].

To these definitions we add:

Definition 2 *Persistent Weak Achievement Goal*

$$\begin{aligned}(\textbf{PWAG}\ x\ y\ p\ q) \stackrel{\triangle}{=} & [\neg(\textbf{BEL}\ x\ p) \wedge (\textbf{PGOAL}\ x\ \Diamond p)] \vee \\ & [(\textbf{BEL}\ x\ p) \wedge (\textbf{PGOAL}\ x\ \Diamond(\textbf{MB}\ x\ y\ p))] \vee \\ & [(\textbf{BEL}\ x\ \Box\neg p) \wedge (\textbf{PGOAL}\ x\ \Diamond(\textbf{MB}\ x\ y\ \Box\neg p))] \vee \\ & [(\textbf{BEL}\ x\ \neg q) \wedge (\textbf{PGOAL}\ x\ \Diamond(\textbf{MB}\ x\ y\ \neg q))]\end{aligned}$$

The **PWAG** requires more commitment from an agent then is required by a **WAG**. Upon discovering that p has been achieved or 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. We will use the **PWAG** during team formation. The following proposition follows directly from the definitions of **WAG** and **PWAG**.

Proposition 1 $(\boldsymbol{PWAG}\ x\ y\ a\ p) \Rightarrow (\boldsymbol{WAG}\ x\ y\ a\ p)$

4 Communicative Acts

Communicative acts will serve as the means agents use to communicate their mental states and to ask for the assistance of other agents in achieving their goals. By performing a particular communicative act, an agent is deliberately attempting to alter the state of the world. In general there is a specific result the act is intended to accomplish — its perlocutionary effect. This result

is related to one or more of the agent's goals. If we assume reliable communications channels, the communications act will succeed in the sense that the addressee will receive it[1]. However, since the communicating agent's goal is to alter the addressee agent's mental state in a particular way, there is no guarantee the act will achieve this ultimate goal. Therefore, in addition to the agent's actual goal, we specify a minimum achievement the act will accomplish — its illocutionary effect. A communicative act is defined as an attempt, which requires four arguments: the agent, the act the agent will perform, the goal of the attempt, and the result the agent has committed to. The two final arguments are the agents ultimate goal and the minimum acceptable achievement. The following formalization of an attempt is from [CL90a].

Definition 3 *Attempt:*
$(\textbf{ATT}\ x\ e\ p\ q) \triangleq (\textbf{BEL}\ x\ \neg p) \wedge (\textbf{GOAL}\ x\ (\textbf{HAPPENS}\ x\ e;p?)) \wedge (\textbf{INT}\ x\ e;q?)$

The agent believes some condition p to not be true at the present time, and is going to do an action (e). The ultimate goal of the agent's act is to bring about some condition p, with the intention of achieving at least q. The agent has only a limited commitment to the ultimate goal (p). In contrast, the agent has an intention to achieve q. 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.

4.1 Request

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

Definition 4 *Request*
$(\textbf{REQ}\ x\ y\ e\ a\ p) \triangleq (\textbf{ATT}\ x\ e\ \phi\ \psi)$
where ϕ *is:*
$\Diamond(\textbf{DONE}\ y\ a) \wedge (\textbf{INT}\ y\ a\ [\textbf{PWAG}\ x\ y\ (\textbf{DONE}\ y\ a)\ p])$
and ψ *is :*
$(\textbf{BMB}\ y\ x\ (\textbf{PWAG}\ x\ y\ [(\textbf{DONE}\ y\ a) \wedge (\textbf{PWAG}\ y\ x\ (\textbf{DONE}\ y\ a)\ (\textbf{PWAG}\ x\ y\ a\ p))]))$

The goal (ϕ) of a request consists of two parts, the first is the straight-forward requirement that the addressee perform the requested act. The second part of the goal places a requirement on the addressee's mental state. The requested action is that y not only perform a, but perform it with respect to x's **PWAG** that y do a (relative to p). It is not enough that y do a, y must also intend to do a relative to x's goal that she[2] do it. 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 request may not fully succeed, because the addressee may refuse the request for any number of reasons. For example, the requested action may conflict (either directly or indirectly) with

[1] Our analysis is valid without the assumption of reliable communications channels. However we would have to account for the possibility of extra messages being sent if there is reason to suspect the original was not received.

[2] We will use she when referring to agent y, and he when referring to agent x.

other of addressee's goals, the addressee may not have the resources to adopt the request due to prior commitments or the addressee may just arbitrarily refuse the request. Under any of these circumstances, the speaker is not obligated to perform the request again.

The minimum result x is committed to is ψ that y believe x has a weak achievement goal that both she eventually do a, and have a weak achievement goal to eventually do a. 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.

Our definition of a team is based on the notion of a **JPG** (joint persistent goal) — the **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, she 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 under no such obligation. The requesting agent is 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 confirm or a refuse.

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.

Axiom 1 *Sincerity Axiom*

($\forall$x ϵ agent, $\forall$e ϵ events)
(**GOAL** x [(**HAPPENS** x e);(**BEL** y p)?]) $\Rightarrow$ (**GOAL** x [(**HAPPENS** x e);(**KNOW** y p)?])

Proposition 2 *(**DONE** (**REQ** x y e a p)) $\Rightarrow$ (**PWAG** x y (**DONE** y a) p)*

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

Proposition 3 *(**DONE** (**REQ** x y e a p)) $\Rightarrow$ (**PGOAL** x (**DONE** y a) p)*

Proof: By Proposition 2 x has a **PWAG** of (**DONE** y a). The definition of a **PWAG** requires x to either have a **PGOAL** of (**DONE** y a), to believe it to be already true or impossible or to believe the relativizing condition p to be false (in the latter three cases x has a **PGOAL** to notify y). Since a **REQ** is an *attempt*, and x is a sincere agent, he cannot believe (**DONE** y a) to be either true or impossible. For the same reason he cannot believe p to be false.

4.2 Refuse

Receipt of a request does not commit the receiving agent to accept it — the agent may refuse a request for any reason, such as a prior conflicting goal. Thus, a refusal is a notification to a requesting agent that this agent will not commit to the goal of the request.

Definition 5 Refuse

*(**REFUSE** y x e a)* $\stackrel{\triangle}{=}$ *(**ATT** y e ϕ ϕ)*
where ϕ *is*
*(**BMB** x y* $\neg$*(**PWAG** y x (**DONE** y a)* ψ*))*
and ψ *is*
*(**PWAG** x y (**DONE** y a) p)*

In the definition of a refusal, the goal and the minimal result are the same. 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 refuse is simply an attempt by the original requestee to make known to the original requestor that she will not commit to the requested action. Our assumption of reliable communication channels and the sincerity axiom imply the attempt will succeed, the act's goal and minimal result are the same.

The definition of a refuse speech act does not imply y cannot undertake a; but only that y will not commit to a with respect to x's **PWAG** . If y were to achieve a by accident or as a by product of some other goal it is pursuing (either independently or in conjunction with other agents), her actions would be consistent with her refusal. While this may, at first, seem to be a minor distinction it has practical consequences. If y were to achieve a after refusing the **PWAG**, she would be under no requirement to notify x, and x might never find out. For x to gain what he wants, a team to form as a result of the request, y would have to accept the **PWAG**.

The main effect of a refusal is to notify the requesting agent that the requested agent cannot be counted upon as a means to achieve the requested action. Thus, a result of a refuse is the requesting agent is freed of the team obligations with respect to the requestee that were incurred by making the original request. The requestor can drop the obligations (the **PGOAL**) that was embedded in her **PWAG**.

Theorem 1

$\vdash$ **HAPPENED** [(**REQ** x y e a p)); (**REFUSE** y x e a p) $\Rightarrow$
$\neg$(**PGOAL** x (**DONE** y a) p)

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, 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.

4.3 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 requesters **PWAG** , the speaker is committed to do a and is also committed to the other obligations of team membership.

Definition 6 *Confirm*

(**CONFIRM** y x e a p) $\stackrel{\triangle}{=}$(**ATT** y e ϕ ϕ)
where ϕ *is*
(**BMB** x y (**PWAG** y x (**DONE** y a) (**WAG** x y (**DONE** y a) p)))

In our definition of **CONFIRM** the ultimate goal and the minimum acceptable goal of the underlying attempt are again the same. As was the case with **REFUSE** we can define an attempt in this manner because of the assumptions we made about the agents and their environments. The existence of reliable communications channels means the message will get to the listener and be understood. Because the speaker is sincere, and the addressee knows it, the addressee will believe the speaker.

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) p)

Proposition 5

$\vdash$ (**DONE** (**CONFIRM** y x e a p)) $\Rightarrow$ (**PGOAL** y (**DONE** y a) p)

I have omitted proofs for these propositions as they are analogous to those of propositions 2 and 3.

4.4 Assert

ASSERT is used by an agent to attempt to establish mutual belief that a particular state of affairs is true.

Definition 7 *Assert*
ASSERT *(y x e q)* $\stackrel{\triangle}{=}$*[**ATT** y e (**BMB** x y q) (**BMB** x y (**BEL** y q))]*

The speaker's goal is for there to be mutual belief that q is true. Performing the attempt does not guarantee this goal will be achieved; for example the addressee could have access to information (that the speaker doesn't have) indicating the speaker is wrong, in this case the addressee will continue to believe that q does not hold. The minimum acceptable result for the speaker is there be mutual belief that the speaker believes that q is true. The speaker is committed to making the addressee believe a fact about his mental state. Our assumption of reliable communication channels guarantees the achievement of this result.

By performing an Assert, the speaker is claiming to believe the truth of q. The definition and the sincerity axiom are enough to prove the following proposition[3]

Proposition 6 $\vdash$*(**HAPPENED** (**ASSERT** y x e q)) $\Rightarrow$ (**BEL** y q)*

[3]In fact, we can define **CONFIRM** and **REFUSE** in terms of **ASSERT**, but space precludes our giving the analysis.

5 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 p, to achieve a. The caveat is necessary because we make an implicit assumption that no event occurs between the the **REQ** and the **CONFIRM** that either makes the goal of the **JPG** true or impossible, or makes the relativizing condition false.

Theorem 2

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

Proof sketch: The **REQ** and the **CONFIRM** each supplied a **PWAG**. The communication of the **PWAG**s are sufficient to form a **WMG**. The existence of the **WMG** and the mutual belief established by the agent's communication allow the formation of a **JPG** and therefore builds a team.

Having created a team, we must supply a method to dissolve it once the **JPG**s goal has been accomplished. Just as team creation is a process that builds interlocking **PWAG**s into a **JPG**, dissolving the team is a process that unwinds the **PWAG**s 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. For instance the assertion that a goal has been achieved, followed by the addressee believing the asserted fact is a sufficient condition for the team members to each drop their **PGOAL** .

Theorem 3

$\vdash$ [**HAPPENED** (**REQ** x y e0 a p); (**CONFIRM** y x e1 a);
e2;(**ASSERT** y x e3 (**DONE** y a));(**BEL** x (**DONE** y a))?] $\Rightarrow$
[$\neg$(**PGOAL** x y a p) $\wedge$ $\neg$(**PGOAL** y x a p)]

The event sequence *e2* in the statement of this theorem represent a sequence of unspecified actions that y has performed or caused to be performed (possibly by recruiting yet other agents) that have achieved the goal a.

It is important to note that although the team members can drop their individual persistent goals, the necessary conditions that will allow the team to disband have still 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 **PWAG**s 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 4

$\vdash$ [**HAPPENED** (**REQ** x y e0 a p); (**CONFIRM** y x e1 a);
e2;(**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))

x's assertion that he agrees that y has **DONE** a is enough to establish mutual belief that the teams goal has been accomplished. Having discharged the **PGOAL** (theorem 3) and established mutual belief, the agents have satisfied the **PWAG**. Mutual belief also satisfies the **JPG**, allowing the agents to drop the associated **WMG**. With this last requirement discharged, the team is dissolved.

6 Application of the Theory

Many researchers have asserted that interagent dialogue follows a finite state model [FFMM94, BDBW95, WF88]. Although we have concerns about the adequacy of this model for human interactions [Coh94], it may be an adequate model for a communications protocol among software agents. Therefore, in this section we will examine the Winograd and Flores' [WF88] model of basic conversation for action, with the intent of showing that the behavior of the transition graph of this model of dialogue is an emergent behavior of agents behaving according to the principles in our model of interagent communication.

The diagram, reproduced as Figure 2, is easy to follow and may indeed describe the basic flow of a task oriented dialogue; however it offers little or no explanation of why the conversation is structured in this way. In the figure 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. We are not claiming these assertions are in error, in fact our claim is that the formalization of speech acts based on joint intention theory can explain the reasons for this classification, and show the transitions in the figure are correct.

Winograd and Flores state the initial request from state 1 to state 2 represents a request stating some *conditions of satisfaction* (their emphasis [WF88, pg. 65]). In our model this is a request of the form (**REQ** A B e p q). A is requesting that B perform some action p with respect to a relativizing condition q. This is a task oriented dialogue, whose purpose is to build a team to accomplish p. A's **REQ** speech act starts that process by informing B that A has a task to be done, and that A wants B to do the task. State 1 serves as the initial state of the dialogue. In this model it is the only initial state. By performing the **REQ**, A has committed resources towards team formation. From 2 and 3 A has outstanding goals. The dialogue must continue until A can discharge those goals. States in which A is able to accomplish this will be the final states of this dialogue.

The remainder of this section will examine the paths through the diagram that start at the initial state and end in some final state. Each of these paths will be analyzed with regard to our inner level

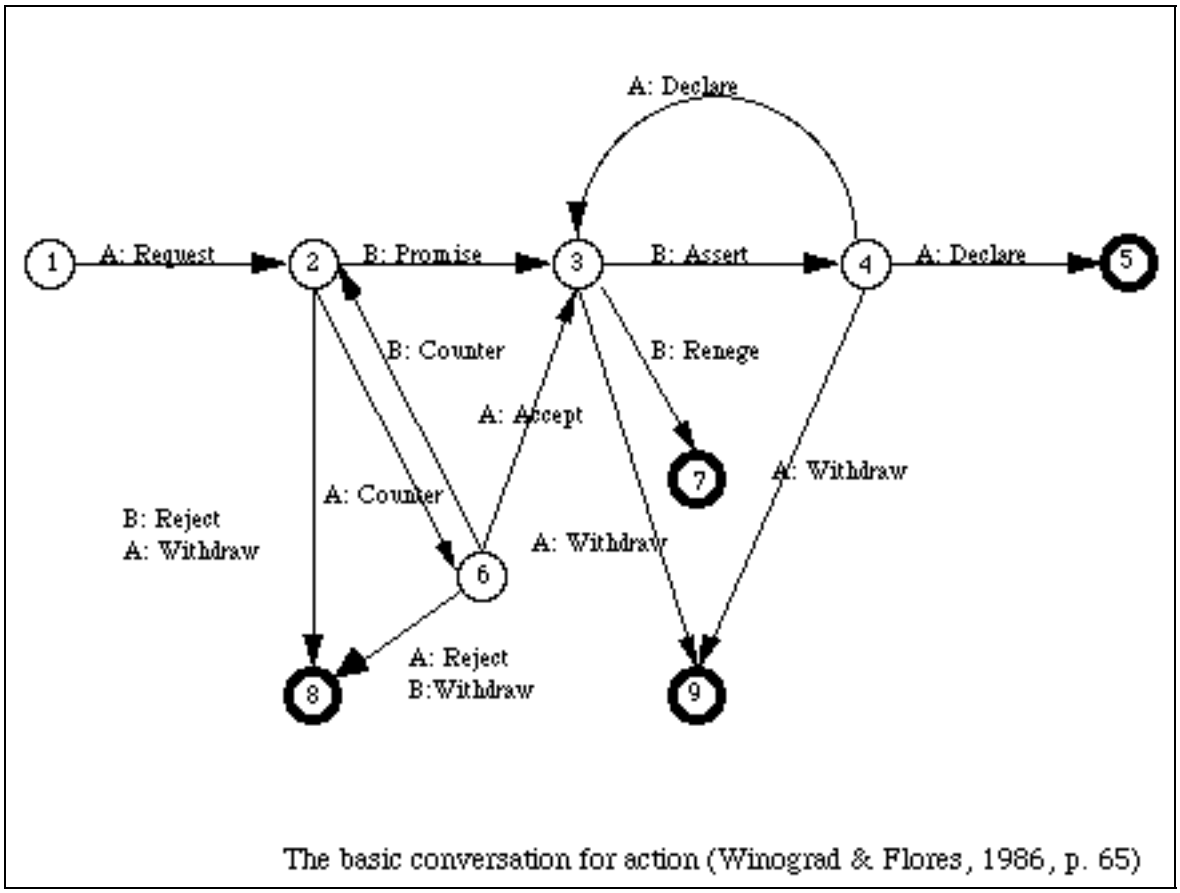
A: Declare
1
A: Request
2
B: Promise
3
B: Assert
4
A: Declare
5
B: Counter
B: Renege
A: Accept
7
A: Withdraw
A: Counter
B: Reject
A: Withdraw
A: Withdraw
6
A: Reject
B:Withdraw
8
9
The basic conversation for action (Winograd & Flores, 1986, p. 65)

Figure 2:

of analysis. In addition we will be able to characterize the type of outer level completion (successful or unsuccessful with regard to the completion of the original task) a conversation following that path achieved.

Starting from the initial state Winograd and Flores claim there are exactly five alternative sets of paths for the conversation to take.

> ... the hearer can accept the conditions ..., can reject them, or can ask to negotiate a change in the conditions of satisfaction (counteroffer). The original speaker can also withdraw the request before a response, or can modify its condition. [WF88, pg. 65]

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 1. The result of this theorem is the requester no longer has any obligations toward potential team members, as is signified by his withdrawal. The hearer's immediate refusal means he never assumed any obligations. An analysis of the other paths leading to state 8 requires a definition of a counteroffer. A counter offer 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 that request.

Definition 8 *Counteroffer*

*(**CO** x y a a1)* $\triangleq$
*(**REF** x y e a);(**ASSERT** x y e1* ϕ*)*
where ϕ *is:*
*(**HAPPENS** (**ASSERT** y x (**WAG** y x (**DONE** x a1) q))) $\Rightarrow$*
*(**INT** 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 and withdrawal allow the application of Theorem 1. The counteroffer illustrates an important point in the design of an agent communication language. There is no need to have a large extendible set of primitive communication acts in the language. When new capabilities are required, they can be built (by composition) from the existing set of communication acts. The reader is referred to [CL95] 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, 8 is a final state. The dialogue has been brought to a successful conclusion. However, on the outer level of dialogue analysis, state 8 represents a failure; 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 2, 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 (**ASSERT** A B (**WAG** A B (**DONE** B a1) q)). *a1* and *q* are bound by B's counteroffer. B's counteroffer followed by A's accept have created a set of interlocking **WAG**s, 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 2 applies. A team has been formed, both A and B have undischarged goals 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 goals.

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 $\Box\neg$(**DONE** B a)). From this we know (**BEL** B $\Box\neg$(**DONE** B a)), and with reliable communication channels A will know this as well. This allows the two parties to discharge the **JPG** and disband the team. Actually there is a requirement that A communicate his acceptance to B before B can completely discharge the **WAG** — in a real world application of the theory, this final communication might take place between the two party's attorneys. In a similar manner 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**. Again, B will have to communicate back to fully dissolve the **JPG**.

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 (**DONE** B a))), that is A is agreeing with B's prior assertion. This represents the situation described in theorem 4. All goals have been discharged, and the team is disbanded with its task accomplished.

7 Conclusions and Future Work

This paper has attempted to show that we can define communications acts in terms of the requisite 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 successfully explains the structure of the protocol. In the process we have demonstrated that our small set of primitive acts can be composed to define more complex communicative acts.

References

[BDBW95] Jeffery Bradshaw, Stewart Dutfield, Pete Benoit, and John D. Wooley. KAoS: Toward an industrial-strength open agent architecture. In J. M. Bradshaw, editor, *Software Agents*. AAAI/MIT Press, 1995.

[CL90a] Philip R. Cohen and Hector J. Levesque. Intention is choice with commitment. *Artificial Intelligence*, 42:213–261, 1990.

[CL90b] Philip R. Cohen and Hector J. Levesque. Rational interaction as the basis for communication. In Cohen et al. [CMP90], chapter 12, pages 221–256.

[CL91] Philip R. Cohen and Hector J. Levesque. Teamwork. *NOÛS*, 21:487–512, 1991.

[CL95] Philip R. Cohen and Hector J. Levesque. Communicative actions for artificial agents. In *Proceedings of the First International Conference on Multi-Agent Systems*. AAAI Press / MIT Press, 1995.

[CMP90] Philip R. Cohen, Jerry Morgan, and Martha E. Pollack, editors. *Intentions in Communication*. System Development Foundation Benchmark Series. Bradford Books, MIT Press, 1990.

[Coh94] Philip R. Cohen. Models of dialogue. In T. Ishiguro, editor, *Cognitive Processing For Voice and Vision*. Society of Industrial and Applied Mathematics, 1994.

[FFMM94] Tim Finin, Richard Fritzson, Don McKay, and Robin McEntire. KQML as an agent communication language. In *Proceedings of the Third International Conference on Information and Knowledge Management*. ACM Press, 1994.

[LCN90] Hector J. Levesque, Philip R. Cohen, and José H. T. Nunes. On acting together. In *Proceedings of the Eighth National Conference on Artificial Intelligence*, pages 94–99. AAAI, AAAI Press, 1990.

[WF88] Terry Winograd and Fernando Flores. *Understanding Computers and Cognition*. Addison-Wesley, 1988.