Designing conversation policies using joint intention theory *

Ira A. Smith † Philip R. Cohen† Jeffrey M. Bradshaw‡§ Mark Greaves‡
Heather Holmback‡

Abstract

In this paper we apply the semantics sketched by Smith and Cohen [11] to the system of conversational policies developed for KAoS [1]. We show that the compositional nature of the speech-acts and their semantics allows us to provide a comprehensive semantics for the conversational policies; and that we can use the generated semantics to analyze policies, to alter policies, and to show how the policies can be combined to form complete dialogues and protocol control strategies. We also show that the semantics provides us with a valuable tool to analyze and to suggest corrections for the conversational policies.

1. Introduction

Research has suggested that conversations among agents can be controlled by finite state networks of speech-acts [1, 13]. In the KAoS system, Bradshaw uses finite state models to construct a set of agent conversation schemas he calls conversation policies. These policies are schemas in that they specify the form of a conversation, but leave open the content. Two policies are illustrated in Figures 1 and 2. Bradshaw and others suggest that these conversation policies can be either the result of emergent behavior (in the case of intelligent agents) or the result of "off-line" design [1, 6]. In either case, the agent must behave in conformance with the obligations incurred by the communication acts used in the policy.

If conversation policies are to be used as the basis for inter-agent communication, they must have precise semantics. Intelligent agents must be able to analyze a policy to determine their behavioral choices at each state in a conversation when the policy is inferred. Unless there is a clear semantics associated with each state, there is no unambiguous way to do this analysis. In the case of less or non-intelligent agents the agent designer must perform the task of analysis, but the requirement for an unambiguous semantics remains.

Conversation policy design also requires an unambiguous assignment of semantics. A designer must be able to prove the correctness of a particular policy. For example, a designer should be able to prove that all the transitions in the policy are possible, that no necessary transitions are missing, and that states labeled as final can actually be the endpoints of a conversation. In addition, policy designers and users intend that a chosen policy be suitable for a particular purpose. To prove that a policy is suitable for use in a particular circumstance we must be able to specify what the expected future behavior of the participating agents will be at each of the final states of the conversation.

It is our contention that a communication act semantics based on the theory of joint intentions provides a basis for deriving a semantics of conversation policies. In this paper we analyze two of the conversation policies suggested by Bradshaw. In the process we show that the joint intention based semantics not only provides an unambiguous meaning for each of the policies, but is also a useful tool to analyze a policy for consistency, fitness for a particular task and to suggest ways policies can be linked together to form more complex dialogues.

Often the purpose of inter-agent conversation is to form a team to accomplish some task. Joint intention theory provides a vehicle to describe the mental states of team participants. Communication acts, when supplied with a semantics based on the theory, are one method to form, maintain and disband teams. For example, as has been previously shown, a conversation consisting of a REQUEST followed by the appropriate CONFIRM places the participants in the men-

*This research was supported in part by the Information Technology Promotion Agency, Japan, as part of the Industrial Science and Technology Frontier Program, "New Models for Software Architecture" sponsored by NEDO (New Energy and Industrial Technology Development Organization), in part by grant N00014-95-1-1164 from the Office of Naval Research, in part by a grant from the Boeing Company to OGI, in part by grant R01 HS09407 from the Agency for Health Care Policy Research to the Fred Hutchinson Cancer Research Center and by the Aviation Extranet joint-sponsored research agreement between NASA Ames, The Boeing Company and the University of West Florida.

†Center for Human Computer Communication, Oregon Graduate Institute

‡Applied Research and Technology, The Boeing Company

§Institute for Human and Machine Cognition, University of West Florida

tal states required for team formation [11]. Our analysis of a revised offer conversation policy proves that a standing offer followed by the appropriate INFORM provides another method to form a team.

In this paper we apply the theory of joint intentions to propose a semantics for a particular set of communication acts, and then use those semantics to develop semantics for selected conversation policies. In section 2 we briefly outline the proposed semantics of communication acts based on the theory of joint intentions. In section 3 we outline selected conversation policies developed in the KAoS system [1], and we apply the tools of joint intention theory to analyze the semantics of the proposed conversation policies.

2. Underlying Semantics

CL theory of intentions

The semantics of the communication acts that are the components of the conversation policies are based on the Cohen-Levesque analysis of joint intentions [3, 4, 5, 10]. Joint intention theory presents a denotational semantics based on the mental states of the agents using these acts to communicate. A basic principle of the semantics is *compositionality* – the semantics of a complex act can be derived from the semantics of the acts that are its syntactic components. Using this model, it is possible to define an agent communication language that allows for principled extensibility. The language can be defined with a small core set of operators; additional operators can be defined in terms of the core set and added to the language. The semantics of the newly added operators are defined to be the compositional result of the component core operators in its definition. This type of semantics stands in contrast to the operational semantics proposed for KQML by Finin and Labrou [8, 9]. Since KQML's operators are defined independently, it is not possible to determine the exact relationship between a pair of operators. Hence neither the language implementor nor the language user is given guidance about how the operators should fit together.

The language used to express the theory is the language of first order logic augmented with modal operators to express propositional attitudes and to talk about sequences of events. The primitive mental states in joint intention theory are an agent's beliefs and goals, expressed as (BEL x p) and (GOAL x p), where *x* is an agent and *p* is a proposition that follows from *x*'s beliefs or goals. A persistent goal, PGOAL, is defined as a goal the agent is committed to keeping until it is either achieved, shown to be impossible or shown to be irrelevant. In addition to these mental states, the theory uses the following operators: (AGT x e) says *x* is the only agent for the (sequence of) events *e*; (HAPPENS a) and (DONE a) say that action expression *a* will happen next or has just happened. AFTER, BEFORE, and LATER are defined using HAPPENS. Knowledge (KNOW) and the usual forms of mutual belief (ABEL, BMB, and MB) are defined in the standard manner. Details of this semantics can be found in [3].

A cornerstone to the theory of joint intentions is the notion of an agent's commitment to achieving some state (in the world), which is expressed as a persistent goal. An intention is a persistent goal in which the agent commits to performing an action while he believes he is about to do it. This analysis has been extended to groups of agents, who can be said to have joint commitments and intentions. The types of organizations that agents form can be characterized by commitment that they have to the group's goal and to the other members of the group. Two important types of joint commitment are a joint persistent goal (JPG), and a persistent weak achievement goal (PWAG). Formal definitions of these states are given in Definitions 1 and 2.

Definition 1 *Joint Persistent Goal*

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

This definition states that two agents (*x* and *y*) have a joint persistent goal of *p*, with respect to *q*, when precisely the following conditions hold: there is mutual belief that *p* does not currently hold, it is a mutual goal that *p* eventually hold, and furthermore that *p* will remain a weak mutual goal (for *x* and *y*) until there is mutual belief that either *p* currently is true, *p* will never be true or *q* is no longer true. A weak mutual goal (WMG) is a mutual belief among the agents that both have weak achievement goals (WAG) with regard to achieving p. A formal definition is presented in [5, 10, 11].

Definition 2 *Persistent Weak Achievement Goal*

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

This definition states that for an agent *x* to have a PWAG with respect to another agent *y* either the agent believes that *p* does not currently hold and has a persistent goal to achieve *p*, or the agent is ready to abandon the persistent goal of *p* for one of three reasons – he believes it to be already true, he believes it to be impossible or he believes the relativizing condition (*q*) no longer holds. In each of the latter three cases, the agent will have the original goal – to achieve *p* – replaced with another persistent goal, that of creating the

mutual belief of the corresponding condition with the other agent.

A PWAG is more than a commitment to perform an action, and more than the commitment to perform an act in concert with another agent. Not only is the agent committed to achieving *p*, the PGOAL in the first disjunct, once *p* is achieved or dropped, the agent will acquire a persistent goal to achieve mutual belief that the goal has been achieved (or is impossible or is no longer relevant). Agents bound together with PWAGs form very resilient groups that we refer to as teams. Teams of agents, once formed, can successfully function under adverse conditions, including breaks in communications and other difficulties in achieving the team's goals [10, 11, 12]. One way that teams, and other organizations of agents, can be formed is through communication acts [11].

Communication-acts as attempts

By performing a communication act, an agent is attempting to influence the mental state of one or more other agents. Communication acts must be characterized as attempts because there is a possibility the act may not fully succeed, that is it may not achieve its ultimate goal. For example if I want the door opened, and ask you to do it (a request), my goal is achieved, if you open the door. However if you do not – you may not want the door open, or it may be locked and you cannot open it – then, although my goal in making the request has not been achieved, I have succeeded in making the request. The following formalization of an attempt is from [4].

Definition 3 *Attempt:*

(**ATT** x e p q) $\triangleq$((**BEL** x $\neg$p) $\wedge$
(**GOAL** x (**HAPPENS** x e;$\Diamond$p?)) $\wedge$
(**INT** x e;q?))?;e

An event *e* is an attempt by agent *x* to achieve an ultimate goal *p*, with a minimum acceptable result of achieving some intended effect *q*, if the agent believes *p* is not currently true, and the agent intends *e* achieve at least *q*.

The distinction between the attempt's goal and intention is very important. If the attempt does not achieve the goal, the attempting agent may retry the attempt, or try some other strategy or even drop the goal. If the attempt does not succeed in achieving the intention, *q*, the agent is committed to retrying. For example if *x* asks *y* to open the door, and *y* refuses, *x* is not obligated to continue asking. However, if *x* believes that *y* didn't hear or understand, he will ask again.

The following definitions of REQUEST and INFORM are derived from the definitions of REQUEST and ASSERT given in [11].

Definition 4 *REQUEST*

(**REQ** x y e a p) $\triangleq$(**ATT** x e ϕ ψ)
where ϕ is:
(**DONE** y a) $\wedge$ (**PWAG** y x [**PWAG** x y (**DONE** y a) p])
and ψ is :
(**BMB** y x (**PWAG** x y [(**DONE** y a) $\wedge$
(**PWAG** y x (**DONE** y a) (**PWAG** x y (**DONE** y a) p))]))

The goal of the request is that the requestee should perform the action in a particular mental state – the requestee must not only intend to perform the action, he must perform it with respect to the requestor's PWAG that he do it. The intention of the act is that the requestee believe that it is mutually believed that the requester has a PWAG that corresponds to the previously described goal of the act.

Often an agent's request will be for the performance of a subtask in an overall plan. In such a case the relativizing clause (*p*) in the requestor's PWAG may be the goal of the plan. It is important to note the level of commitment that the goal of a REQUEST requires of the requestee – the requestor's goal is not simply that *y do a*, but that, in addition, *y* have a PWAG to do *a* with respect to *x*'s PWAG that *y* do *a*.

The semantics for REQUEST also requires a high level of commitment from the requesting agent, because his PWAG requires him to notify the other agent should he change his mind or discover a problem with the REQUEST. By communicating his PWAG the agent is *already* treating the agent receiving the REQUEST as a team member [11]. This meaning of REQUEST differs from the semantics of the REQUEST in the FIPA ACL specification [6], for which the feasibility preconditions are based only on the intentions of the requesting agent. While this is sufficient to communicate the content of the REQUEST, it is not a strong enough commitment to begin forming a team. For example, should the requesting agent drop the intention behind the REQUEST, nothing in the FIPA definition would require him to notify the other agent.

Definition 5 *INFORM*

(**INF** y x e q) $\triangleq$[**ATT** y e (**BMB** x y q) (**BMB** x y (**BEL** y q))]

The goal of an INFORM is that there be mutual belief, between the listening agent (*x*) and the informing agent (*y*), that the proposition *q* is true. The intention of the INFORM is that there be mutual belief that the listening agent believe the informing agent believes *q*.

Smith and Cohen have previously defined the CONFIRM and REFUSE communication acts as specialized types of INFORM. They demonstrated that a REQUEST followed by the appropriate CONFIRM is sufficient to create the interlocking PWAGs, and hence the JPG, required to build a team [11]. The formal definitions of these communication acts follow.

Definition 6 *CONFIRM*

(**CONFIRM** x y a p) $\stackrel{\triangle}{=}$(**INF** x y a ϕ)
where ϕ is:
(**PWAG** x y (**DONE** x a) (**PWAG** y x (**DONE** x a) p))

Definition 7 *REFUSE*

(**REFUSE** x y a p) $\stackrel{\triangle}{=}$(**INF** x y ϕ)
where ϕ is
($\Box\neg$(**PWAG** x y (**DONE** x a) (**PWAG** y x (**DONE** x a) p)))

By using an CONFIRM, *x* is informing agent *y* that he has a PWAG (to *y*) to perform *a* with respect to *y*'s PWAG (that *x* do *a*). A REFUSE has the opposite effect, *x* is informing *y* that he does not have (in fact, that he never will have) such a PWAG. Note that while *x* has informed *y* that he does not have the PWAG, there is nothing in the meaning of REFUSE that would prevent *x* from doing *a*.

3. Conversation policies

Conversations are sequences of messages involving two or more agents intended to achieve a particular purpose. We analyze conversations through the use of *conversation policies*, which are schema that can be used to define and restrict what can take place in individual agent conversations. In the past, we have formulated these schema as finite state machines with transition arcs signifying different allowable speech acts, although there is no reason why other more expressive formalisms could not be used. The key ideas here are: 1) Conversation policies provide a level of analysis that abstracts from the precise propositional content, agent communication language, and implementation of individual conversations; and 2) These policies (which have been constructed offline) can help ensure reliable communication among heterogeneous agents while lessening the burden of inferring which communication acts are possible for the next message in a conversational sequence.

Conversation policies can simplify the design of agent communications modules because they are considerably less complicated to implement than unrestricted agent dialogue models. However, we are still left with the question of how to design, analyze, verify, and test proposed conversation policies. In this section, we will use our semantics of communication acts [11] as a tool to analyze two of the conversation policies proposed in [1], the $\mathrm{Inform}_{\mathrm{CP}}$ and the $\mathrm{Offer}_{\mathrm{CP}}$ conversation policies. For the purpose of this analysis, we will make certain simplifying assumptions about the interacting agents: the agents are sincere and each believes the other to be sincere; the transport mechanism is reliable and each agent believes the communication transport mechanism to be reliable; that agents know at all times what conversation policy they are currently using and which state they are in, and they share a common ontology and bind the same symbols to objects in the world.

Inform Conversation Policy

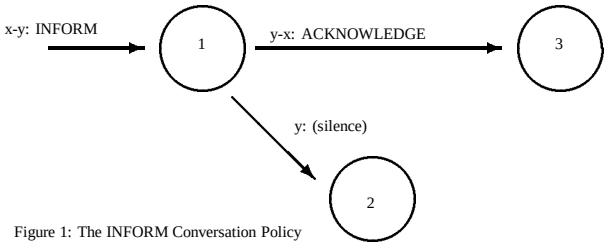

Figure 1: The INFORM Conversation Policy

The inform conversational policy (Figure 1) can be used by agent *x* to send a message (*p*) to agent *y*. From the definition of an INFORM, we know something about agent *x* 's mental state after the transition to state 1 of the policy. The requirements of the inform are that *x*'s goal be that *y* believe *p* to be true and that *x*'s intention is that *y* believe that *x* believes *p*. We can (using our assumptions) also conclude that *x*'s INFORM will achieve, at least, its intention. Since *x*'s intention has been satisfied, he is left with no unsatisfied commitments as the result of his inform. Please note that we cannot conclude that *y* will believe *p* – although *y* will believe that *x* believes *p*, *y* may have information (that presumably *x* does not have) that causes him to doubt or outright not believe *p*.

There are two arcs leading from state 1. The arc leading to state 3 represents *y* acknowledging *x*'s INFORM. We agree with Bradshaw's [1] characterization of the tension between economy of verb types (e.g. communication acts) and naturalness of expression, and we suggest a definition of ACKNOWLEDGE as a specialization of INFORM.

Definition 8 *ACKNOWLEDGE*

(**ACKNOWLEDGE** x y e p c) $\stackrel{\triangle}{=}$
c?;(**INF** x y e (**BEL** x (**BEL** y p)))
where c is:
$\exists$e1.(**DONE** (**INF** y x e1 p))

An ACKNOWLEDGE must take place in the context of a prior INFORM, which was provided by the original transition to state 1. The ACKNOWLEDGE will not create any new commitments for either *x* or *y*, so we can conclude that when the conversation policy reaches state 3, no commitments have been created for either conversant.

States 1 and 2 are connected by an arc representing silence, that is no communicative action by *y*. Once again no commitments have been created for either *x* or *y*. We may conclude that states 2 and 3 represent final states of a conversation policy, in that, any commitments created by

the communication acts in the policy have been discharged. These states are final states only with respect to the policy – if the policy were to be embedded in a larger dialogue, these states may not be final states of the dialogue.

Offer Conversation Policy

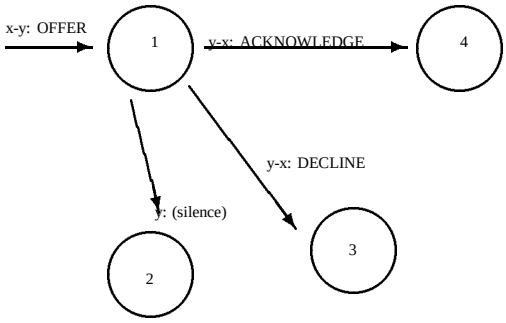

Figure 2: The OFFER Conversation Policy

The offer conversation policy – $\mathrm{Offer_{CP}}$ (as illustrated in Figure 2) is under-specified, allowing for at least two possible interpretations. The first interpretation is one in which the ACKNOWLEDGE is to be regarded as acknowledging nothing more than receipt. Since neither the ACKNOWLEDGE nor the silence transition would convey acceptance of the OFFER, the policy would require the offering agent to act on his offer unless he receives a DECLINE from the agent receiving the OFFER. This is analogous to the situation where a bystander notices someone precariously laden with packages trying desperately to open a door and says 'Let me help you'. Without waiting for a verbal response, the offeror quickly reaches for the door handle and opens it. Given this understanding of $\mathrm{Offer_{CP}}$, an OFFER with the following semantics could initiate such a policy:

Definition 9 *OFFER*

(**OFFER** x y e a) $\stackrel{\triangle}{=}$
(**INF** x y e [**PGOAL** x (**DONE** x a)
(**BEL** x (**GOAL** y [$\Diamond$**DONE** a]))])

Agent *x* is informing agent *y* that he has a persistent goal to perform *a*, this persistent goal being contingent upon *x*'s belief that *y* wants *a* done. Unlike commitments based on joint intentions or PWAGs, the commitment implied by such an OFFER is fragile. *X*'s persistent goal (to do *a*) can be dropped at any time if *x* comes to believe that *y* does not have a goal that *a* be done. Certainly the transition to state 3 will provide that belief. However, if *x* comes to the belief in any other way, he does not have a commitment to inform *y*. Furthermore, *x* has no commitment to inform *y* even if he succeeds in doing *a*. Unless *y* can directly sense the result of *a*, there is no guarantee that *y* will ever find out if *x* has succeeded or failed. In addition, *x* cannot be sure that *y* wants *a* to be done. If *y* changes his goals, *x* cannot expect *y* to inform him. This stands in contrast to *x* and *y*'s forming a team. Members of a team are bound by interlocking PWAGS requiring any team member who has dropped the goal (in this case that *a* be done) to be left with the goal of achieving mutual belief with other team members.

Because of the fragility of the connection between agents who are cooperating in this fashion, this interpretation of $\mathrm{Offer_{CP}}$ is worthwhile in only very limited circumstances. Agents should be reluctant to interact in this manner when the action offered would require the expenditure of significant resources or time. In an environment where the action requires an extended time or the expenditure of significant resources, an agent making the offer will want assurance that the recipient wants the act done and continues to want it done throughout the time it takes to perform. The recipient will want to be informed if the offering agent finishes the promised act or if he abandons the attempt. This leads to an alternative interpretation for the conversation policy.

In the second interpretation of $\mathrm{Offer_{CP}}$, the ACKNOWLEDGE must be replaced with an INFORM containing an acceptance of the OFFER, and the silence must be coupled with a timeout. Under this interpretation, silence can be regarded as equivalent to a DECLINE, thereby allowing agents who may be deluged with 'junk-mail' offers some alternative besides being forced to make explicit responses to each one. This is the interpretation implemented in the KAoS system.

The altered policy is illustrated in Figure 3. The altered policy begins with a *standing offer* – SOFFER, a formal definition of which follows:

Definition 10 *Standing Offer (**SOFFER**)*

(**SOFFER** x y e a) $\stackrel{\triangle}{=}$(**INF** x y e ϕ)
where ϕ is:
$\forall$e1.(**DONE** y (**INF** y x e1
(**PWAG** y x (**DONE** x a)))) $\Rightarrow$
(**DONE** e1;(**PWAG** x y (**DONE** x a)
(**PWAG** y x (**DONE** x a))))

Y is making a conditional offer – *if you will inform me that you have a PWAG that I do* a, *I will have a PWAG to do* a *(relative to your PWAG)*. The agent making the offer is protecting himself – he will not acquire a commitment unless the recipient confirms that he wants *a* done. In addition, should the recipient of the offer change his goals, the PWAG requires him to insure that the offering agent becomes aware of the change. The standing offer also protects the recipient– the offering agent's PWAG guarantees the recipient will be told in the event of either success or failure.

The semantics of our SOFFER are related to those of KQML's *advertise* performative and to the FIPA ACL's *propose* [6, 9]. In all three cases, the agent is commiting to a

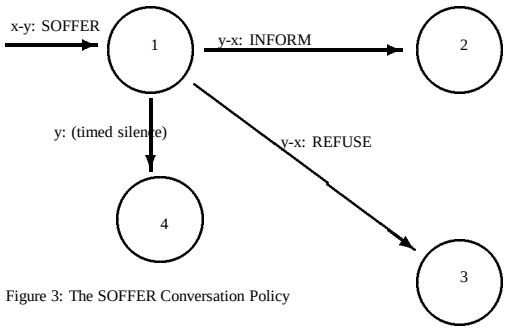

Figure 3: The SOFFER Conversation Policy

future action (if asked). For SOFFER, the commitment extends over time – the offering agent is agreeing to perform the action if the listening agent *ever* makes it known that he has the appropriate PWAG. FIPA's *propose* requires the agent receiving the *propose* to currently have the appropriate intention. KQML's *advertise* represents a larger promise than the commitment contained in our SOFFER – the advertising agent is agreeing to perform the action anytime any agent asks. The scope of this pledge points out the need for a KQML agent to be able to inform other agents of its capabilities without acquiring commitments towards its future actions.

Since the only transitions from state 1 in this standing offer conversation policy require a response from the recipient agent, the offering agent will not end in an ambiguous mental state. A REFUSE will inform agent *x* that he has no PWAG. Because a REFUSE does not engender a commitment by the recipient, state 3 is a final state of the policy. With the interpretation of timed silence as a REFUSE, the same argument will apply to state 4. An INFORM by the recipient having the propositional content specified in the standing offer definition will create PWAGs between the offering and the recipient agents. By the definition of a team in [11], these PWAGs are sufficient to form a team with (DONE y a) as its motivating goal. State 2 is a final state of this conversation policy, but the PWAGs created leave both agents with unresolved commitments that form a team. The PWAGs require further actions – the offering agent must perform the action offered, and further communication must occur to disband the team [11]. So, while state 2 may be a final state of the conversation policy, it cannot be a final state of a complete dialogue. The nature of the agent's PWAGs requires there to be further interaction to discharge the commitments of both agents.

A standing offer is an INFORM of a conditional commitment in which the speaker is informing the listener that he (the speaker) will acquire a commitment to perform an action (*a*), with respect to the listener's goal to have *a* done, if she (the listener) informs him that that is her goal. Below, we show that if the original listener performs the INFORM, not only will the original speaker's PWAG will be triggered but a team will be formed.

Theorem 1 *An* SOFFER *by x to y to perform a, followed by an* INFORM *by y to x that he wants x to do a and a CONFIRM by x is sufficient to form a team with respect to the propositional content of the* SOFFER. *In formal terms:*

(**DONE** [(**SOFFER** x y a);
(**INF** y x e1(**PWAG** y x (**DONE** x a)));
(**CONFIRM** x y (**PWAG** x y (**DONE** x a)))]) ⇒
(**JPG** x y (**DONE** x a) (**PWAG** y x (**DONE** x a)))

Proof sketch:

1. The offering agent (*x*) has made a conditional offer (to *y*). Since *x* is a sincere agent, by the definition of an SOFFER, *x* will have a PWAG toward *y* to do *a* if *y* informs *x* that she has a PWAG that *x* do *a*.

2. Agent *y* has a PWAG that *x* do *a*, because *y* is sincere. She had the PWAG when she performed *e1*. Based on the definition of a PWAG *y* either still wants *x* to do *a*, or is communicating to *x* that she no longer wants *a* done.

3. The assumption of a reliable communication channel guarantees that *x* heard *y*'s inform. Because *x* believes *y* is sincere, after *y*'s INFORM *x* believes that *y* has a PWAG that *x* do *a*.

4. After *y*'s inform, by steps 1 and 3, *x* has a PWAG to do *a* relative to *y*'s goal.

5. Because *x* is sincere and the communication channel is reliable, after *x*'s CONFIRM *y* believes *x* has the PWAG.

6. Because both agents are sincere, and each believes the other to be sincere, and because the communication channels are reliable and known to be reliable, steps 1, 3 ,4 and 5 are sufficient to establish mutual belief in each others PWAGs.

7. By steps 2, 4, 5, 6 and the definition of a joint persistent goal, *x* and *y* have a JPG that *x* do *a* (relative to *y*'s goal).

8. Based on the definition of a team, and on step 7, *x* and *y* form a team with the task that *x* do *a* relative to *y*'s goal.

It is instructive to contrast the formation of a team in this manner with the method of team formation described by Smith and Cohen [11]. When a team is formed by a REQUEST followed by a CONFIRM, the initiating agent is making a commitment to treat the other agent as a team member, at least for the purposes of communication, even before receiving the listener's reply. If the requestor changes his goal, he has the obligation to inform the other agent of the change. The listening agent can depend on this and knows

the outcome of the exchange will depend on his response. When a team is formed with an SOFFER followed by the appropriate INFORM, neither agent can be sure of the outcome until the end of the conversation. The offering agent does not know if his offer will be accepted and while the informing agent is considering the offer, he does not know if the offer still holds. Nothing in the definition of an SOFFER requires the offering agent to notify the other agent if the offer is no longer valid – until the SOFFER is accepted. These characteristics of the $\mathrm{Soffer}_{\mathrm{CP}}$ conversation policy makes team formation using this policy more fragile than in the case of a policy that uses a REQUEST and a CONFIRM, but it does follow our intuitive notion of what constitutes an offer. Offers made between people have an implicit temporal impermanence. If certain resources are required to perform *a*, the offering agent is not obligated to reserve them if a response is not immediately received. So if there is a long enough delay between the SOFFER and a positive response, the offering agent may not be able to perform *a*, even though the SOFFER was sincere at the time it was made. Although the agent accepting the SOFFER cannot be sure the offer is still in effect when he accepts it, if conditions have changed and the SOFFER is no longer effective, the accepting agent can expect the offering agent to inform him. In this case, where the SOFFER is no longer in force, the offering agent's PWAG will still come into effect, and since he can no longer do *a*, the agent would be obligated (by the definition of the PWAG contained in the SOFFER) to inform the other agent that the action was no longer possible.

If we closely examine the context and the content of the INFORM speech act, we can prove that the utterance has an effect that implies a REQUEST has been done. The goal of a REQUEST is to have the listener perform an action (*a*) while in the mental state where the listener is intending to do *a* because the speaker wants it done. An INFORM is also defined as an attempt, which has a goal for the listener to believe the propositional content of the INFORM. In the context of the standing offer conversation policy, the agent performing the INFORM (*y*) already believes the offering agent (*x*) is willing to adopt a PWAG to do *a* with respect to *y*'s goal, and that he will do so if informed that *y* has such a PWAG. The propositional content of the INFORM is that agent *y* has a PWAG that agent *x* do *a*, with respect to *y*'s goal that he do so. Given that *y* has the PWAG and is sincere, *y* must also have this as a goal, and is making this goal public with the INFORM. This goal is the same goal as a REQUEST from *y* to *x* that he do *a*. The intention of that REQUEST is that there be mutual belief that *x* believe that *y* has this PWAG, which is precisely the intention of the INFORM, and leads us to another theorem.

Theorem 2 *In the context of an open standing offer by an agent x to perform an action a, an inform to x by y that he has a PWAG that x do a implies that a REQUEST by y that x do a has been done. In formal terms this is expressed as:*

$\forall$e1,e2. (**DONE** (**SOFFER** x y e1 a);
(**INF** y x e2 (**PWAG** y x (**DONE** x a)))) $\Rightarrow$
(**DONE** (**SOFFER** x y e1 a);(**REQ** y x e2 (**DONE** x a)))

Proof sketch:

1. For a speech act to be a REQUEST certain conditions must be met. First the speaker (*y*) must be in the appropriate mental state, namely he must have the proper PWAG: in this case, (PWAG y x (DONE x a)). This is the same PWAG required for the SOFFER, hence this requirement is met.

2. A second requirement is that the speech act has the correct intention. In our case this requires the intention to be that agent *x* believe it is mutually believed that *y* has a PWAG for *x* to do *a* and to be in the correct mental state – namely that *x* have a PWAG to do *a* with respect to *y*'s PWAG. The act *e2*, coupled with reliable communication channels and *y*'s belief in *x*'s sincerity, ensures the INFORM will be heard and believed by *x*, and that this is known to *y*. Since both *x* and *y* believe *y* has the PWAG, and each is aware of the other's belief, it can be shown that after the INFORMs both *x* and *y* believe it is mutually believed that *y* has the correct PWAG [11]. Hence the intention requirement is met.

3. The final requirement is that the goal of the speech act be for the listener (*x*) to eventually do *a* in the correct mental state – a PWAG to do *a* with respect to the speaker's (*x*) PWAG. This is the consequent of *x*'s original SOFFER. The content of *y*'s INFORM meets the requirements of the antecedent of the SOFFER. By performing the event *e2*, *y* is expressing the goal that *x* believe that she has the goal (PWAG x y (DONE x a)). Hence *x* must have the goal that *y* have the proper PWAG.

4. Conclusion and Future work

In this paper we have shown how the theory of joint intentions can be used as a tool for the analysis of conversation policies. We used joint commitments to analyze the KAoS $\mathrm{Inform}_{\mathrm{CP}}$ and $\mathrm{Offer}_{\mathrm{CP}}$ conversation policies with respect to the mental states of the participants. Our analysis exposed certain ambiguities in the $\mathrm{Offer}_{\mathrm{CP}}$ policy. We then used the theory to propose a definition for the SOFFER speech act and to design a conversation policy. We also proved that an SOFFER followed by the appropriate INFORM will form a team, and that in the context of an SOFFER, the respondent's INFORM implies that a REQUEST speech act has been done.

The mental states of the participants of the first interpretation of the $\mathrm{Offer}_{\mathrm{CP}}$ conversation policy suggests that agents can organize themselves for cooperation in ways other than in teams. These modes of organization might be fragile and even disintegrate in situations that teams could survive, yet they could also require far less overhead and effort to maintain. In situations where the agents are in direct perceptual view of each other, or in cases when the consequences of failure are minimal, these fragile organizations might suffice. Research remains to be done to analyze the behavioral characteristics of these organizations using joint intention theory.

Researchers in the area of agent communication languages and multi-agent communication have worked under the assumption that many of the problems of natural language discourse do not apply to agent communication languages. The basis for this assumption is that the artificial language and the relatively simple, constrained conversations preclude these complications. Our observation that the INFORM in $\mathrm{Soffer}_{\mathrm{CP}}$ implies that a REQUEST has been done shows that given a content language rich enough to allow the embedding of modal operators, the analysis of multi-agent conversation may need to account for indirect speech acts.

In the past, we have used finite state machines to represent the allowable speech act sequences in our conversation policies. However, finite state machines are only one possible way to formally represent constraints on speech act sequences. More expressive formalisms can be constructed out of statements in a suitable dynamic logic; less expressive (but perhaps more readily understandable) formalisms could be built out of regular expression grammars. One could also imagine a conversation policy consisting of a higher-level set of constraints which, rather than specifying the exact sequence and type of speech acts involved, allowed for some degree of flexibility. Such a policy, for example, might describe a relative sequence or pattern of "landmarks" in a conversation of a given type, each landmark defining a set of specifiable semantic properties that must hold of the agents involved (e.g. an offer has been made; an offer has been accepted), and the overall policy simply requiring that the transitions between conversational landmarks be made by an appropriate sequence of one or more speech acts. Policies of this type provide a pathway to truly *emergent* agent conversations, in which agents perform explicit planning over the messages they produce in order to achieve their communication goals.

We intend to continue our analysis of conversation policies to help answer these and other questions that are sure to arise. Over the long term we would like to develop tools that would help agent conversation designers be sure that the policies they are creating are both consistent with their assumptions about what the conversations are designed to accomplish and with the underlying semantic model of commitments [2, 7].

References

[1] J. M. Bradshaw et al. KAoS: Toward an industrial-strength open agent architecture. In J. M. Bradshaw, editor, *Software Agents*. AAAI/MIT Press, 1997.

[2] J. M. Bradshaw et al. Mediating representations for an agent design toolkit. In *Proceedings of the Eleventh Annual Knowledge Acquisition Workshop (KAW(98))*, 1998.

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

[4] P. R. Cohen and H. Levesque. Rational interaction as the basis for communication. In P. R. Cohen, J. Morgan, and M. E. Pollack, editors, *Intentions in Communication*. MIT Press, 1990.

[5] P. R. Cohen and H. J. Levesque. Teamwork. *Nous*, 21:487–512, 1991.

[6] Foundation for Intelligent Physical Agents. Fipa 97 specification part 2: Agent communication language. http://drogo.cselt.it/fipa, 1997.

[7] M. Greaves, H. Holmback, and J. M. Bradshaw. CDT: A tool for agent conversation design. In *AAAI-98 Workshop on Software Tools for Developing Agents*. AAAI, 1998.

[8] Y. Labrou. *Semantics for an Agent Communication Language*. PhD thesis, University of Maryland, Baltimore County (UMBC), 1996.

[9] Y. Labrou and T. Finin. A proposal for a new KQML specification. www.cs.umbc.edu/kqml/kqmlspec.ps.

[10] H. J. Levesque, P. R. Cohen, and J. Nunes. On acting together. In *Proceedings of the National Conference on Artificial Intelligence*. AAAI Press, 1990.

[11] I. A. Smith and P. R. Cohen. Toward a semantics for an agent communication language based on speech-acts. In *Proceedings of the National Conference on Artificial Intelligence*. AAAI Press, 1996.

[12] M. Tambe. Agent architectures for flexible, practical teamwork. In *Proceedings of the National Conference on Artificial Intelligence*. AAAI Press, 1997.

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