A Social Semantics for Agent Communication Languages

Munindar P. Singh[*]

Department of Computer Science
North Carolina State University
Raleigh, NC 27695-7534, USA

singh@ncsu.edu

Abstract. The ability to communicate is one of the salient properties of agents. Although a number of agent communication languages (ACLs) have been developed, obtaining a suitable formal semantics for ACLs remains one of the greatest challenges of multiagent systems theory. Previous semantics have largely been mentalistic in their orientation and are based solely on the beliefs and intentions of the participating agents. Such semantics are not suitable for most multiagent applications, which involve autonomous and heterogeneous agents, whose beliefs and intentions cannot be uniformly determined. Accordingly, we present a social semantics for ACLs that gives primacy to the interactions among the agents. Our semantics is based on social commitments and is developed in temporal logic. This semantics, because of its public orientation, is essential to providing a rigorous basis for multiagent protocols.

1 Introduction

Interaction among agents is the distinguishing property of multiagent systems. Communications are a kind of interaction that respect the heterogeneity and preserve the autonomy of agents. In this respect, they differ from physical interactions. An agent may have no choice but to physically affect another agent—e.g., to bump into it or to lock a file it needs—or similarly be affected by another agent. By contrast, unless otherwise constrained, an agent need not send or receive communications; if it is willing to handle the consequences, it can maintain silence and deny the requests or even the commands it receives. Consequently, communication is unique among the kinds of actions agents may perform and the interactions in which they may participate.

Our particular interest is in *open* multiagent systems, which find natural usage in modern applications such as electronic commerce. In open multiagent systems, the member agents are contributed by several sources and serve different interests. Thus, these agents must be treated as

- *autonomous*—with few constraints on behavior, reflecting the independence of their users, and
- *heterogeneous*—with few constraints on construction, reflecting the independence of their designers.

Openness means that all interfaces in the system, and specifically ACLs, be given a clear semantics. A good ACL semantics must meet some crucial criteria.

- *Formal.* The usual benefits of formal semantics apply here, especially (1) clarity of specifications to guide implementers and (2) assurance of software. In fact, these are more significant for ACLs, because ACLs are meant to be realized in different agents implemented by different vendors.
- *Declarative.* The semantics should be declarative describing what rather than how. Such a semantics can be more easily applied to a variety of settings not just those that satisfy some low-level operational criteria.
- *Verifiable.* It should be possible to determine whether an agent is acting according to the given semantics.
- *Meaningful.* The semantics should be based on some intuitive appreciation of communications and not treat communications merely as arbitrary tokens to be ordered in some way. If it does, we can arbitrarily invent more tokens; there would be no basis to limit the proliferation of tokens.

These criteria, although simple, eliminate all of the existing candidates for ACL semantics. For example, English descriptions of communications (quite common in practice) are not formal, finite-state machines (FSMs) are not declarative or meaningful, mentalistic approaches are not verifiable, and temporal logics (if applied to tokens directly) and formal grammar representations of sequences of tokens are not meaningful. Briefly, we find that an approach based on social commitments (described within) and formalized in temporal logic can meet all of the above requirements.

Organization The rest of this paper is organized as follows. Section 2 motivates a semantics based on social constructs, and presents our conceptual approach. Section 3 presents a formal social semantics for ACLs, discusses its properties, and shows how it relates to communication protocols. Section 4 concludes with a discussion of our major themes, the literature, and some questions for future investigation.

2 Conceptual Approach

Most studies of communication in AI are based on speech act theory [1]. The main idea of speech act theory, namely, to treat communications as actions, remains attractive. An *illocution* is the core component of a communication and corresponds to what the communication might be designed to (or is meant to) accomplish independent of both how the communication is physically carried out (the *locution*) and the effect it has on a listener (the *perlocution*). For example,

I could request you to open the window (the *request* is the illocution) by saying so directly or hinting at it (these are possible locutions). Whether or not you accede to my request is the perlocution. A proposition can be combined with illocutions of different types to yield different messages. For example, my request to open the window is different from my assertion that the window is open.

It is customary to classify ACL primitives or message types into a small number of categories based on the different types of illocution. Usually these include the following categories—a sample primitive of each category is given in parentheses: *assertives* (*inform*), *directives* (*request*), *commissives* (*promise*), *permissives* (*permit*), *prohibitives* (*forbid*), *declaratives* (*declare*), and *expressives* (*wish*). The above classification provides sufficient structure for our approach (some alternative classifications could also be used). Each message is thus identified by its sender and receiver, (propositional) content, and type.

Three components of an ACL are typically distinguished: (1) a content sublanguage to encode domain-specific propositions, (2) a set of primitives or message types corresponding to different illocutionary types (e.g., *inform* and *request*), and (3) a transport mechanism to send the messages. Part (2) is the core and the most interesting for the present study.

2.1 Mentalistic versus Social Semantics

Work on speech act theory within AI was motivated from the natural language understanding perspective and concerned itself with identifying or inferring the "intent" of the speaker. As a result, most previous work on ACLs too concerns itself with mental concepts, such as the beliefs and intentions of the participating agents. In fact, some theories emphasize the mutual beliefs and joint intentions of the agents as they perform their communicative actions. Inferring the beliefs and intentions of participants is essential to determining whether a given illocution occurred: did the speaker make a signal or was he just exercising his arm? However, in ACLs, the message type is explicit and no reasoning is required to determine it. In applications of multiagent systems, such reasoning would usually not be acceptable, because of the difficulty in specifying, executing, and enforcing it in an open environment.

There are a number of objections to using only the mental concepts for specifying ACL semantics; several of these are described in [18]. Although the mental concepts might be suitable for specifying the construction and behavior of the agents, they are not suitable as an exclusive basis for communications. There are a number of objections, but we summarize them in the following major categories.

– *Philosophical.* Communication is a public phenomenon, but the mental concepts are private. Any semantics that neglects the public nature of communication is deeply unsatisfactory. Something obviously takes place when agents interact through language even if they don't have or share the "right" beliefs and intentions.

– *Practical.* Ensuring that only the desirable interactions occur is one of the most challenging aspects of multiagent system engineering. However, the mental concepts cannot be verified without access to the internal construction of the agents. Under current theories of mental concepts, we cannot uniquely determine an agent beliefs and intentions even if we know the details of its construction.

The above evidence supports the conclusion that a purely mentalistic semantics of an ACL cannot be a normative requirement on agents or their designers.

2.2   Language versus Protocol

To ensure autonomy and heterogeneity, we must specify communications flexibly and without making intrusive demands on agent behavior or design. Traditionally, heterogeneity is accommodated by specifying communication protocols. Traditionally, the protocols are specified by defining the allowed orders in which communicative acts may take place, but no more. Often, this is through the use of FSMs. In particular, FSM protocols are devoid of content. They only state how the tokens are ordered. Thus, the ACL is effectively discarded, and we can just as well choose any arbitrary tokens for the message types. The same holds for other formalisms such as push-down automata, formal grammars, Petri Nets, and temporal logic (when these are applied on tokens), so we won't discuss them explicitly here.

The foregoing indicates how the unsuitability of the traditional semantics forces the protocols to be ad hoc. By contrast, the present paper seeks to develop a nontrivial semantics for an ACL that would also be usable for the construction and verification of protocols.

2.3   Validity Claims

The semantics of ACLs, which concerns us here, relates to the essence of communication. The currently popular approaches to ACL semantics are based on the speaker's intent [8]. Under this doctrine, the illocution is what the speaker believed and intended it to be. This doctrine, championed by Searle and others, however, leads to the philosophical and practical problems discussed above.

In philosophy, another of the best known approaches to communicative action is due to Habermas [9]; short tutorials on Habermas are available in [13] and [23, chap. 2]. The Habermas approach associates three "worlds" or aspects of meaning with communication. These correspond to the three *validity claims* implicitly made with each communication:

– *objective*, that the communication is true.
– *subjective*, that the communication is sincere—in other words, the speaker believes or intends what is communicated.
– *practical*, that the speaker is justified in making the communication.

In conversation, each of the above claims may be challenged and shown to be false. However, even if false, these claims are staked with each communication, which is why they can be meaningfully questioned. The claims involve different aspects of meaning including the subjective, but by fact of being claims in a conversation, they are public and social. If I tell you something, I am committed to being accurate, and you are entitled to check if I am. I am also committed to being sincere, even though you may not be able to detect my insincerity unless you can infer what I believe, e.g., through contradictory statements that I make at about the same time. In general, in open environments, agents cannot safely determine whether or not another agent is sincere.

Perhaps more than his followers in AI, Searle too recognizes the institutional nature of language. He argues that the "counts as" relation is the basis for "constitutive reality" or institutional facts, including definitions of linguistic symbols [15, pp. 152–156] and [16, chap. 4]. But institutions are inherently objective. For example, in an auction, raising your hand counts as making a bid whether or not you have the intention to actually convey that you are bidding. In on-line commerce, pushing the "submit" on your browser counts as authorizing a charge on your credit card irrespective of your intentions and beliefs at that time.

Our proposed approach, then, is simply as follows. We begin with the concept of *social commitments* as is studied in multiagent systems [3] and reasoning and dialogue in general [24]. Our technical definition of commitments differs from the above works in two main respects [19, 22]. Our formalization of commitments includes

- the notion of a social context in the definition of a commitment; the social context refers to the team in which the given agents participate and within which they communicate; it too can be treated as an agent in its own right— e.g., it may enter into commitments with other agents.
- metacommitments to capture a variety of social and legal relations.

The different claims associated with a communicative action are mapped to different commitments among the participants and their social context. Consequently, although our semantics is social in orientation, it admits the mental viewpoint.

Social commitments as defined are a kind of deontic concept. They can be viewed as a generalization of traditional obligations as studied in deontic logic. Traditional obligations just state what an agent is obliged to do. In some recent work, directed obligations have also been studied that are relativized to another agent—i.e., an agent is obliged to do something for another agent. Social commitments in our formulation are relativized to two agents: one the beneficiary or creditor of the given commitment and another the context within which the commitment occurs. Further, we define operations on commitments so they can be created and canceled (and otherwise manipulated). The operations on commitments are, however, subject to metacommitments.

In our approach, metacommitments are used to define micro societies within which the agents function. These are intuitively similar to the *institutions* of

[14], which however also specify the meanings of the terms used in the given (trading) community.

3 Technical Approach

Communication occurs during the execution of a multiagent system. For this reason, our semantics is based on commitments expressed in a logic of time.

3.1 Background Concepts

Temporal logics provide a well-understood means of specifying behaviors of concurrent processes, and have been applied in areas such as distributed computing. By using classical techniques, such as temporal logic, we hope to facilitate the application of the proposed semantics when multiagent systems are to be integrated into traditional software systems. Computation Tree Logic (CTL) is a branching-time logic that is particularly natural for expressing properties of systems that may evolve in more than one possible way [5]. Conventionally, a model of CTL is a tree whose nodes correspond to the states of the system being considered. The branches or *paths* of the tree indicate the possible ways in which the system's state may evolve.

Our formal language $\mathcal{L}$ is based on CTL. $\mathcal{L}$ builds on a flexible and powerful variety of social commitments, which are the commitments of one agent to another. A commitment involves three agents: the *debtor* (who makes it), the *creditor* (to whom it is made), and the *context* (the containing multiagent system in the scope of which it is made). We include beliefs and intentions as modal operators.

The following Backus-Naur Form (BNF) grammar with a distinguished start symbol L gives the syntax of $\mathcal{L}$. $\mathcal{L}$ is based on a set Φ of atomic propositions. Below, *slant* typeface indicates nonterminals; $\longrightarrow$ and | are metasymbols of BNF specification; $\ll$ and $\gg$ delimit comments; the remaining symbols are terminals. As is customary in formal semantics, we are only concerned with abstract syntax.

L1. $L \longrightarrow$ *Prop* $\ll$atomic propositions, i.e., in $\Phi\gg$
L2. $L \longrightarrow$ $\neg\ L$ $\ll$negation$\gg$
L3. $L \longrightarrow$ $L \wedge L$ $\ll$conjunction$\gg$
L4. $L \longrightarrow$ $L \rightsquigarrow L$ $\ll$strict implication$\gg$
L5. $L \longrightarrow$ $\mathsf{A}\ P$ $\ll$universal quantification on paths$\gg$
L6. $L \longrightarrow$ $\mathsf{E}\ P$ $\ll$existential quantification on paths$\gg$
L7. $L \longrightarrow$ $\mathsf{R}\ P$ $\ll$selecting the real path$\gg$
L8. $P \longrightarrow$ $L \cup L$ $\ll$until: operator on a single path$\gg$
L9. $L \longrightarrow$ $\mathsf{C}(Agent, Agent, Agent, L)\ll$commitment$\gg$
L10. $L \longrightarrow$ $x\mathsf{B}\ L \mid x\mathsf{I}\ L$ $\ll$belief and intention$\gg$

The meanings of formulas generated from L are given relative to a model and a state in the model. The meanings of formulas generated from P are given relative

to a path and a state on the path. The boolean operators are standard. Useful abbreviations include $\mathsf{false} \equiv (p \wedge \neg p)$, for any $p \in \Phi$, $\mathsf{true} \equiv \neg\mathsf{false}$, $p \vee q \equiv \neg p \wedge \neg q$ and $p \rightarrow q \equiv \neg p \vee q$. The temporal operators A and E are quantifiers over paths. Informally, $p\mathsf{U}q$ means that on a given path from the given state, q will eventually hold and p will hold until q holds. $\mathsf{F}q$ means "eventually q" and abbreviates $\mathsf{true}\mathsf{U}q$. $\mathsf{G}q$ means "always q" and abbreviates $\neg\mathsf{F}\neg q$. Therefore, $\mathsf{EF}p$ p will hold on some path. R selects the real path. $\mathsf{RF}p$ means that p will hold on the real path. Although agents can't predict the future, they can make (possibly false) assertions or promises about it.

$M = \langle \mathbf{S}, <, \approx, \mathbf{N}, \mathbf{R}, \mathbf{A}, \mathbf{B}, \mathbf{I}, \mathbf{C} \rangle$ is a formal model for $\mathcal{L}$. $\mathbf{S}$ is a set of states; $< \subseteq S \times S$ is a partial order indicating branching time, $\approx \subseteq S \times S$ relates states to similar states, and $\mathbf{N} : \mathbf{S} \mapsto 2^{\Phi}$ is an interpretation, which tells us which atomic propositions are true in a given state. $\mathbf{P}$ is the set of paths derived from $<$. $\mathbf{PP}$ gives the powerset of $\mathbf{P}$. For $t \in \mathbf{S}$, $\mathbf{P}_t$ is the set of paths emanating from t. $\mathbf{R} : \mathbf{S} \mapsto \mathbf{P}$ gives the real path emanating from a state. $\mathbf{A}$ is a set of agents. $\mathbf{B} : \mathbf{S} \times \mathbf{A} \mapsto \mathbf{S}$, $\mathbf{I} : \mathbf{S} \times \mathbf{A} \mapsto \mathbf{PP}$, and $\mathbf{C} : \mathbf{S} \times \mathbf{A} \times \mathbf{A} \times \mathbf{A} \mapsto \mathbf{PP}$ give the modal accessibility relations for beliefs, intentions, and commitments, respectively.

For p derived from L, $M \models_t p$ expresses "M satisfies p at t" and for p derived from P, $M \models_{P,t} p$ expresses "M satisfies p at t along path P."

M1. $M \models_t \psi$ iff $\psi \in \mathbf{N}(t)$, where $\psi \in \Phi$

M2. $M \models_t p \wedge q$ iff $M \models_t p$ and $M \models_t q$

M3. $M \models_t \neg p$ iff $M \not\models_t p$

M4. $M \models_t p \rightsquigarrow q$ iff $M \models_t p$ and $(\forall t' : M \models_{t'} p \Rightarrow (\forall t'' : t' \approx t'' \Rightarrow M \models_{t''} q))$

M5. $M \models_t \mathsf{A}p$ iff $(\forall P : P \in \mathbf{P}_t \Rightarrow M \models_{P,t} p)$

M6. $M \models_t \mathsf{E}p$ iff $(\exists P : P \in \mathbf{P}_t$ and $M \models_{P,t} p)$

M7. $M \models_t \mathsf{R}p$ iff $M \models_{\mathbf{R}_t,t} p$

M8. $M \models_t x\mathsf{I}p$ iff $(\forall P : P \in \mathbf{I}(x,t) \Rightarrow M \models_{P,t} p)$

M9. $M \models_t x\mathsf{B}p$ iff $(\forall t' : t' \in \mathbf{B}(x,t) \Rightarrow M \models_{t'} p)$

M10. $M \models_t \mathsf{C}(x,y,G,p)$ iff $(\forall P : P \in \mathbf{C}(x,y,G,t) \Rightarrow M \models_{P,t} p)$

M11. $M \models_{P,t} p\mathsf{U}q$ iff $(\exists t' : t \leq t'$ and $M \models_{P,t'} q$ and $(\forall t'' : t \leq t'' \leq t' \Rightarrow M \models_{P,t''} p))$

3.2 Social Semantics

We now present a social semantics for the ACL primitives. Our main purpose with this semantics is to show how the different validity claims can be understood in terms of social commitments and formalized in our framework.

In giving this semantics, we attempt to understand each communication atomically, i.e., as an individual transmission. Clearly communications usually occur in extended protocols. In a strict reading, Habermas too would be against the idea of seeking to understand communications in isolation. However, from a technical standpoint it is simpler if we can characterize the communications individually. Then we can go back to composing them, so that we might, for example, have an explicit acceptance after a request. In a sense, such an acknowledgement is needed to ensure that the receiver becomes committed to

carrying out the request. If we are operating in a social context where the receiver's commitment is given, then explicit acceptance is superfluous. We return to this point in analyzing Winograd & Flores' conversation for action protocol in Section 3.2.

Illocution	Objective	Subjective
$inform(x,y,p)$	$\mathsf{C}(x,y,p)$	$\mathsf{C}(x,y,x\mathsf{B}p)$
$request(x,y,p)$	$\mathsf{C}(y,x,\mathsf{RF}p)$	$\mathsf{C}(y,x,y\mathsf{IF}p)$
$promise(x,y,p)$	$\mathsf{C}(x,y,\mathsf{RF}p)$	$\mathsf{C}(x,y,x\mathsf{IF}p)$
$permit(x,y,p)$	$\mathsf{C}(x,y,\mathsf{EF}p)$	$\mathsf{C}(x,y,\neg x\mathsf{I}\neg\mathsf{F}p)$
$forbid(x,y,p)$	$\mathsf{C}(y,x,\neg\mathsf{RF}p)$	$\mathsf{C}(y,x,\neg y\mathsf{IF}p)$
$declare(x,y,p)$	$\mathsf{C}(x,y,p)$	$\mathsf{C}(x,y,x\mathsf{I}p)$

Table 1. Social semantics formalized: objective and subjective

Illocution	Practical
$inform(x,y,p)$	$\mathsf{C}(x,G,inform(x,y,p)\rightsquigarrow p)$
$request(x,y,p)$	$\mathsf{C}(x,G,request(x,y,p)\rightsquigarrow \mathsf{AFC}(y,x,p))$
$promise(x,y,p)$	$\mathsf{C}(x,G,promise(x,y,p)\rightsquigarrow \mathsf{RF}p)$
$permit(x,y,p)$	$\mathsf{C}(x,G,permit(x,y,p)\rightsquigarrow \neg\mathsf{C}(y,G,\neg\mathsf{RF}p))$
$forbid(x,y,p)$	$\mathsf{C}(x,G,forbid(x,y,p)\rightsquigarrow \mathsf{C}(y,G,\neg\mathsf{RF}p))$
$declare(x,y,p)$	$\mathsf{C}(x,G,declare(x,y,p)\rightsquigarrow p)$

Table 2. Social semantics formalized: practical

Tables 1 and 2 gives the formal semantics of the ACL primitives. (All commitments are relative to G, the context group, which is not shown to reduce clutter; however, G is the creditor of some commitments, which are shown.) This semantics simply captures the objective, subjective, and practical meanings associated with the given primitive. Each aspect of meaning is viewed from the public perspective, because each involves a social commitment. Let's consider each component of the semantics in turn.

Objectively, the sender commits for *inform* that its content is true, for *promise* that its content will be accomplished, for *permit* that its content may be realized, for *declare* that its content is true. For *request*, the sender expects that the receiver will commit to making it true, and for *forbid* that the receiver will commit that its content will not be realized. Although these are not part of the objective meaning, they are related to the practical meaning given below.

Subjectively, the sender commits for *inform* that he believes its content, for *promise* that he intends to carry it out, for *permit* that he does not intend the negation of its content, for *declare* that he intends to bring it about. For *request*,

the sender expects that the receiver will commit to intending to make it true, and for *forbid* that the receiver will commit that its content will not be realized. These expectations are not directly incorporated in the semantics.

The practical aspect of the semantics is the most complex. Practically, the sender commits for *inform* that he has reason to know the content, for *promise* that if he promises something he can make it happen, for *permit* that he has the authority to relieve the receiver of any commitment to do otherwise, and for *declare* that his saying so, brings it about. For *request*, the sender commits that the receiver has committed to accepting a request from him. For *forbid*, the sender commits he can cause the receiver to take on a commitment to not let the condition come about. This semantics reflects our intuition that prohibitives such as *forbid* are to be differentiated from directives such as *request*. The requester only has to be committed to the claim that his request will eventually be serviced, whereas the forbidder has to be committed to the claim that his prohibition will immediately commit the receiver to not violating it. The above meanings are naturally phrased as metacommitments to the group. They refer to the communication itself.

Conceivably, even the commitments relating to the subjective expectations might be added here, but we suggest they would be too strong for the basic practical meaning. This is because our goal with this semantics is to specify the objective and the practical components of the semantics for use in the construction and validation of multiagent protocols. This is facilitated when the subjective criteria are not included in the practical meaning.

Notice that any commitment may in principle be broken. However, the breaking of a commitment is typically constrained by some metacommitment, which might prescribe an alternative commitment.

Pragmatic Constraints What we usually refer to informally as *meaning* is a combination of the semantics and pragmatics. We will treat the semantics as the part of meaning that is relatively fixed and minimal. Pragmatics is the component of meaning that is context-sensitive and depends on both the application and the social structure within which it is applied.

The above semantic validity claims, even the practical claims, are different from pragmatics. Pragmatic claims would be based on considerations such as the Gricean maxims of manner, quality, and quantity [7]. For example, a pragmatic claim basis for *permit* might be that the receiver desires or intends the content that is being permitted. Some of the pragmatic constraints would be the public versions of the expectations listed above in the subjective component of the semantics.

Protocols and Compliance The limitations of traditional ACL semantics force protocol approaches to fend for themselves and give low-level, procedural characterizations of interactions. Representations based on monolithic finite-state machines are suitable only for the most trivial scenarios. They cannot accommodate distributed execution, compliance testing, or exceptions. However,

given a commitment-based semantics for ACLs, an observer of a multiagent system (possibly itself a participating agent) can maintain a record of the commitments being created and modified. From these, the observer can determine the compliance of other agents with respect to the given protocol. This compliance testing would be based on the contents of the messages and the formal public meanings of the ACL primitives used. It would not depend solely on the sequence of events in the system.

However, protocols will continue to be useful even when a social semantics for ACLs is adopted. For example, turn-taking might underlie the specific commitments to ensure that they are created only when they make sense. For instance, a bidder shouldn't make a bid prior to the advertisement, or the commitment content of the bid won't be fully defined. Moreover, protocols supply the requirements through which the communications can be composed. In other words, although the commitment-based semantics can tell us the result of composing some communications, it is the protocols that tell us what composition is appropriate.
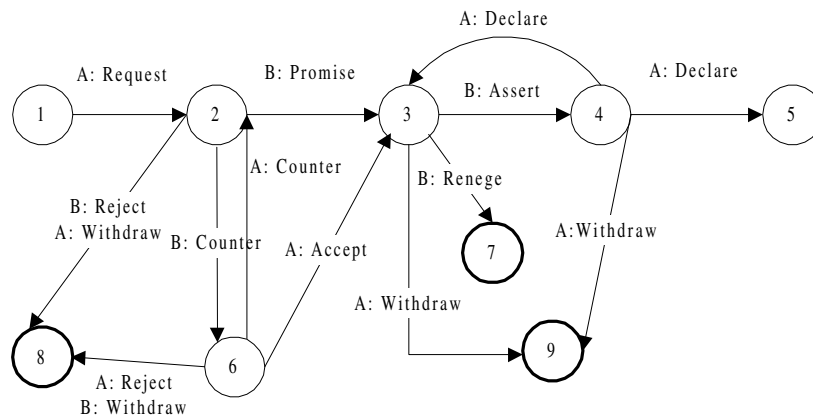

Fig. 1. Conversation for Action [25]

As a simple example, consider the well-known "conversation for action" protocol of Winograd & Flores, which is shown in Figure 1. Commitments can be associated with each state in this protocol. The commitments arise from the basic communication semantics but are enhanced through the metacommitments that are in force in the given organization. This helps us analyze the protocol. For example, if there is a metacommitment that A's requests will be honored, then there is no need to separate states 2 and 3, and in fact, state 6 is eliminated entirely. Conversely, if B makes a promise to A without any explicit request from A, in terms of the commitments, we can see that the protocol effectively begins from state 6. Thus if the applicable metacommitments are captured, the executions are minimally constrained only to satisfy those metacommitments.

This we believe is a major advantage of the declarative approach over low-level representations.

As an aside, we observe that as traditionally stated, this protocol is overspecified, because it mixes commitment concerns with coordination requirements. For example, if no metacommitment applies under which A's request must be honored by B, then the purpose of the transition from state 2 to state 8 only helps in announcing that the protocol has terminated, not in changing any commitments among the participants.

3.3 Properties

A formal semantics may be evaluated by showing that it supports the desirable properties. Our semantics satisfies the four criteria of Section 1.

- *Formal.* Our semantics is based on logic.
- *Declarative.* Our semantics involves assertions about commitments, rather than procedures or automata.
- *Verifiable.* Our semantics offers different levels of verifiability. Every commitment to a putative fact can be verified or falsified by challenging that putative fact. Every commitment to a mental state can be similarly verified or falsified, but only through the more arduous route of eliciting the agent's beliefs and intentions. These might be elicited by observing the agent's further communications or other actions. Of course, this elicitation cannot be reliably performed unless significant assumptions about the agent's design and behavior can be made. As a result, the subjective meaning cannot be used in open environments. Mentalist approaches fail because they almost exclusively consider subjective meaning, although the details of their proposals can vary.

  Every commitment to some institutional fact can be verified or falsified by appeal to some external authority. This authority is the context within which the commitments are created. The context essentially defines the institution within which the communication takes place. The context could be defined as just the group of everyone involved, but in distributed computing practice would refer to some sort of a leader, possibly one that was elected.
- *Meaningful.* Every message type has an inherent meaning expressed in terms of commitments, and arbitrary tokens would be rejected.

We consider some additional technical properties. By being based on the commitments of the participating agents, our semantics provides a basis for describing the conversational state of a multiagent system in high-level terms, i.e., using commitments. Thus, the state, defined in terms of commitments, is independent of the *history*, i.e., the steps of a protocol that may have been executed. History-freedom is essential to establishing some important properties of multiagent protocols, which we discuss next.

- *Composition.* Protocols may be combined into larger protocols, as long as one protocol yields a commitment state that the other protocol needs.

– *Digression.* Ad hoc actions, e.g., in response to exceptions or errors, may be interposed without affecting the true meaning of a protocol as long as the commitments were not affected. If the commitments are affected, then we would know there was a fundamental deviation from the protocol, and may repair it or discard the protocol altogether.
– *Optimization.* The agents may directly enter a protocol in execution where the right commitments are defined even though the steps have not been carried out explicitly. Such short-circuiting of the protocols is crucial for optimizing the agents' behavior.

The above properties are essential for enabling opportunistic behavior by the agents while providing all the benefits of protocols in structuring their interactions and individual behavior. Achieving both flexibility and structure is essential for many multiagent applications.

Similarly, the *content* of a message is not only the direct action it connotes, but also the implied actions caused by the discharge of the applicable metacommitments. Thus, if A has a metacommitment that it will honor B's bid, then B's bid will create the commitment to honor it.

4   Discussion

A communication protocol involves the exchange of messages with a streamlined set of tokens. Traditionally, these tokens are not given any meaning except through reference to the beliefs or intentions of the communicating agents. By contrast, our approach assigns *public*, i.e., observable, meanings in terms of social commitments. This leads to the ability to test compliance at a level of abstraction higher than just the ordering of events. It also promises a canonical form of communication protocols, which would give us a meaningful basis for determining where in a protocol execution the agents in a multiagent system are, how to proceed, and how to accommodate exceptions naturally.

4.1   Literature

Social commitments are not to be confused with commitments previously studied in AI. Traditional commitments apply to an agent being in a state where it will persist with a belief or an intention. They do not reflect the agent's social commitments or obligations to other parties. The notion of persistence with a goal provides a basis for the theory of joint intentions due to [11]. Roughly, a joint intention among some agents corresponds to the agents believing that they have persistent goals to achieve the given condition and that they would inform the others if the condition were to be satisfied or become unsatisfiable or if they drop out of the joint intention for any reason. Thus, joint intentions build on mutual beliefs. Roughly, a set of agents mutually believe p iff each of them believes p, and each of them believes that each of them believes p, and so on, *ad infinitum* [6]. In fact, mutual beliefs are used primarily to establish impossibility results

for distributed computing protocols. Such results can readily be created for joint intentions as well. Traditional approaches typically assume that mutual beliefs can be achieved easily, sometimes by as little as a single message transmission [21, p. 163]. This is unsatisfactory, because a fairly complex theory is built only to be discarded at the first opportunity where it might be tested.

A number of approaches consider the deontic notion of obligation. Traditional obligations involve what a single agent is obliged to do irrespective of other agents. Traum uses such obligations to state what an agent may be obliged to do in a conversation [21]. Traditional obligations, however, do not capture the subtleties of interactions among agents. More promising are directed obligations of the sort studied by Dignum & van Linder [4]. Dignum & van Linder model agents with an explicit social component (the other components are less relevant here). They model speech acts as affecting the beliefs or obligations of the agents. The classification of speech acts in this approach is a little different from the typical classifications. However, their speech acts can be mapped to the more common kinds of message types. Although some of their constructs are similar to ours, Dignum & van Linder confine themselves to giving the preconditions for various messages, rather than keeping the semantics sensitive to the context of usage as we have sought to do. For example, they assume that the agents will be sincere.

Besides the works referred to in the above, there is a fairly substantial body of literature on ACLs and their semantics. The Foundation for Intelligent Physical Agents (FIPA) has been standardizing an ACL along with a formal semantics. This ACL and its semantics are based on Arcol, which was part of a system for human-computer interaction [2]. Arcol and the FIPA ACL are mentalist in their orientation. FIPA also includes interaction protocols, which are characterized purely operationally. Labrou & Finin present a variant of the knowledge query and manipulation language (KQML) [10]. They offer a semantics stating how the beliefs and intentions of the participants are affected by communications.

Smith & Cohen present an alternative semantics for an ACL, which is based on a theory of joint intentions [20]. The approach treats communication as creating or modifying structures of joint intentions, which are used to describe the agents working as a team. The joint intentions treatment of teams, however, is a mentalist approach to simulate an essentially social phenomenon. It fails to describe teams directly and suffers from all the problems attendant to the mentalist approaches.

A number of approaches have studied communication protocols. Usually, these specify and execute protocols in a representation such as FSMs and Petri Nets. Labrou & Finin present a grammar for constructing conversations or protocols. The grammar is fundamentally of the same style of representation as FSMs, but is more expressive [10]. Smith & Cohen apply their approach on the conversation for action protocol described above [20]. They argue that the different paths in the protocol result in the formation or nonformation of a team. Interestingly they consider the paths in Figure 1 and not the states as we sought

to do. Also, because their basic assumptions are strong, they cannot suggest how the protocol may be improved in a nontrivial way.

Singh proposes a semantics that gives the conditions for the "whole-hearted" satisfaction of different communications [17]. Whole-hearted satisfaction depends on the intentions and know-how of the participants. This approach has ingredients relating to the objective and subjective aspects of meaning as described above. However, although whole-hearted satisfaction involved a public stance, it lacks a social perspective as developed here. This approach can state communication constraints for use as inputs into the design of the participating agents.

4.2 Directions

The social semantics developed here treats the social construction of communication as a first class notion rather than as a derivative of the mentalist concepts. Although some previous researchers have discussed the social aspects of communication, they were never quite able to shed the mentalist bias of traditional AI. We hope that by making a fresh start on the semantics, we will be able to produce a semantics that can serve the needs of agents operating and communicating in open environments. Because this work is still in an early stage, some of the details are quite likely to evolve. One of our tasks is to evaluate variations of the above semantics that preserve the social, validity-based theme that we tried to capture above.

Interesting theoretical questions are opened up by the present approach. One of these is the longstanding topic of presuppositions and consensus. *Presuppositions* are essential to understanding and properly interpreting any communication. Potentially, we can interpret the implicit claims behind every communication as presuppositions [24]. If they are not challenged, they become accepted as *consensus*, which corresponds to commitments by the entire group of communicating agents. Consensus might offer a tractable alternative to mutual beliefs, which are used by current theories of dialogue, but which cannot be obtained in realistic environments, e.g., those with unreliable asynchronous communication [6].

References

1. John L. Austin. *How to Do Things with Words.* Clarendon Press, Oxford, 1962.
2. Phillipe Breiter and M. David Sadek. A rational agent as a kernel of a cooperative dialogue system: Implementing a logical theory of interaction. In *ECAI-96 Workshop on Agent Theories, Architectures, and Languages*, pages 261–276. Springer-Verlag, 1996.
3. Rosaria Conte and Cristiano Castelfranchi. *Cognitive and Social Action.* UCL Press, London, 1995.
4. Frank Dignum and Bernd van Linder. Modelling social agents: Communication as action. In *Intelligent Agents III: Agent Theories, Architectures, and Languages*, pages 205–218. Springer-Verlag, 1997.

5. E. Allen Emerson. Temporal and modal logic. In Jan van Leeuwen, editor, *Handbook of Theoretical Computer Science*, volume B, pages 995–1072. North-Holland, Amsterdam, 1990.

6. Ronald Fagin, Joseph Y. Halpern, Yoram Moses, and Moshe Y. Vardi. *Reasoning About Knowledge*. MIT Press, Cambridge, MA, 1995.

7. H. P. Grice. Logic and conversation. In P. Cole and J. L. Morgan, editors, *Syntax and Semantics, Volume 3*. Academic Press, New York, 1975. Reprinted in [12].

8. Paul Grice. Utterer's meaning and intentions. *Philosophical Review*, 1969. Reprinted in [12].

9. Jürgen Habermas. *The Theory of Communicative Action, volumes 1 and 2*. Polity Press, Cambridge, UK, 1984.

10. Yannis Labrou and Tim Finin. Semantics and conversations for an agent communication language. In *Proceedings of the International Joint Conference on Artificial Intelligence*, 1997.

11. Hector J. Levesque, Philip R. Cohen, and Jose T. Nunes. On acting together. In *Proceedings of the National Conference on Artificial Intelligence*, pages 94–99, 1990.

12. Aloysius P. Martinich, editor. *The Philosophy of Language*. Oxford University Press, New York, 1985.

13. Gerald Midgley. The ideal of unity and the practice of plurality in systems science. In Robert L. Flood and Norma R. A. Romm, editors, *Critical Systems Thinking: Current Research and Practice*, chapter 2, pages 25–36. Plenum Press, New York, 1996.

14. Juan A. Rodríguez-Aguilar, Francisco J. Martín, Pablo Noriega, Pere Garcia, and Carles Sierra. Towards a test-bed for trading agents in electronic auction markets. *AI Communications*, 11(1):5–19, 1998.

15. John R. Searle. *The Construction of Social Reality*. Free Press, New York, 1995.

16. John R. Searle. *Mind, Language, and Society: Philosophy in the Real World*. Basic Books, New York, 1998.

17. Munindar P. Singh. *Multiagent Systems: A Theoretical Framework for Intentions, Know-How, and Communications*. Springer-Verlag, Heidelberg, 1994.

18. Munindar P. Singh. Agent communication languages: Rethinking the principles. *IEEE Computer*, 31(12):40–47, December 1998.

19. Munindar P. Singh. An ontology for commitments in multiagent systems: Toward a unification of normative concepts. *Artificial Intelligence and Law*, 1999. In press.

20. Ira A. Smith and Philip R. Cohen. Toward a semantics for an agent communications language based on speech-acts. In *Proceedings of the National Conference on Artificial Intelligence*, pages 24–31, 1996.

21. David R. Traum. A reactive-deliberative model of dialogue agency. In *Intelligent Agents III: Agent Theories, Architectures, and Languages*, pages 157–171. Springer-Verlag, 1997.

22. Mahadevan Venkatraman and Munindar P. Singh. Verifying compliance with commitment protocols: Enabling open web-based multiagent systems. *Autonomous Agents and Multi-Agent Systems*, 2(3):217–236, September 1999.

23. Egon M. Verharen. *A Language-Action Perspective on the Design of Cooperative Information Agents*. Catholic University, Tilburg, Holland, 1997.

24. Douglas N. Walton and Erik C. W. Krabbe. *Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning*. State University of New York Press, Albany, 1995.

25. Terry Winograd and Fernando Flores. *Understanding Computers and Cognition: A New Foundation for Design*. Addison-Wesley, Reading, MA, 1987.