In Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI), 1991

Towards a Formal Theory of Communication for Multiagent Systems

Munindar P. Singh*

Dept of Computer Sciences and Artificial Intelligence Lab
University of Texas MCC
Austin, TX 78712-1188 Austin, TX 78759
USA USA

Abstract

Agents in multiagent systems interact to a large extent by communicating. Such communication may be fruitfully studied from the point of view of speech act theory. In order for multiagent systems to be formally and rigorously designed and analyzed, a semantics of speech acts that gives their objective model-theoretic conditions of satisfaction is needed. However, most research into multiagent systems that deals with communication provides only informal descriptions of the different message types used. And this problem is not addressed at all by traditional speech act theory or by AI research into discourse understanding. I provide a formal semantics for the major kinds of speech acts at a level that has not been considered before. The resulting theory applies uniformly to a wide range of multiagent systems. Some applications of this theory are outlined, and some of its theorems listed.

1 Introduction

Multiagent systems are ubiquitous in Artificial Intelligence. Even in the simplest such systems, the epithet is justified only if the agents involved interact with each other in different ways. One of the most natural ways in which intelligent interaction may occur is through communication, especially communication about action. Agents may command, request, advise, or permit each other to do certain actions. They may also promise actions of their own, or prohibit those of others. When complex multiagent systems are to be designed or analyzed a formal theory of the kinds of communication that may take place among agents would be crucial. Unfortunately, no theory is currently available that provides the *objective semantics* of the messages exchanged. This paper describes research that has been done to fill this void. This research uses ideas about the ability and intentions of situated agents that were motivated and developed on independent grounds, albeit with a view to their final application to this problem [Singh, 1991; Singh, 1990a; Singh, 1990b]. This connection to other theories is reason to be reassured that this theory is not *ad hoc*, and will coherently fit in a bigger picture. The theory presented in this paper has ramifications in several subareas of AI, notably, multiagent planning and action, autonomous agents, and cooperative work [Gasser and Huhns, 1989; Huhns, 1987].

Traditionally, speech act theory classifies communications or messages into several kinds of *illocutionary acts* [Searle, 1969; Searle and Vanderveken, 1985]. These include *assertives, directives, commisives, permissives* and *prohibitives*. Briefly, assertives are statements of fact; directives are commands, requests or advice; commisives (e.g., promises) commit the speaker to a course of action; permissives issue permissions; and prohibitives take them away. Classical logic applies only to the case of assertives and considers only their truth and falsity. Therefore, it is inappropriate for other kinds of speech acts (Hamblin describes and criticizes several nonclassical logics for commands [Hamblin, 1987, pp. 97–136], so I do not consider them here). Research in speech act theory, on the other hand, concentrates on describing the conditions under which a particular speech act (of whatever form) may be said to have occurred [Grice, 1969; Searle, 1975]. The AI literature in this area too is concerned with the linguistic or discourse-related aspects of this problem (e.g., for identifying the illocutionary force of indirect speech acts [Allen and Perrault, 1980], or defining their effects on the mutual beliefs of agents [Cohen and Levesque, 1988]).

Of interest here is the orthogonal problem of formally describing the conditions of satisfaction for the different kinds of speech acts. I take the view that communication occurs because agents need to interact effectively and to influence each others' actions. While the illocutionary force of a speech act can be trivially determined from the syntax (in an artificial language that our agents would use), the objective conditions of the satisfaction of a speech act are a part of the semantics. A formal semantics is important because (1) as designers and analyzers, we need a rigorous understanding of communication in the systems we design; and (2) we would often like to embed a version of the semantics in the agents themselves so they can use it in their reasoning about their own (and others') speech acts.

*This research was partially supported by the Microelectronics and Computer Technology Corporation, and by the National Science Foundation (through grant # IRI-8945845 to the Center for Cognitive Science, University of Texas).

The main original contributions of this paper are the following: (1) It argues that there is a level of formal semantics of speech acts that is distinct from both (a) what is traditionally considered their semantics, namely, the conditions under which they may be said to have occurred; and (b) their pragmatics, namely, the effects they may or ought to have on the speaker's and hearer's cognitive states. I.e., the proposed semantics differs from both the illocutionary and the perlocutionary aspects of speech acts. (2) This paper argues that the semantics of speech acts roughly corresponds to the conditions under which we would affirm that the given speech act had been satisfied. (3) It is proposed that this semantics can be captured in the usual model-theoretic framework by introducing different operators that distinguish the satisfaction of a speech act from its mere occurrence. (4) The actual definitions are to be given in terms of the intentions and know-how of the participants and the state of the world (at some salient time or times).

A problem not addressed here concerns the effects a speech act has on the hearer. These depend on issues like the social relationship of the agents or on matters of performance—these are not easy to describe, and are connected to processes of deliberation and belief revision [Perrault, 1987], rather than to the semantics of communication *per se*. Perrault provides some postulates for such revision using default logic. His focus is on the pragmatics of speech acts in natural language understanding, rather than the semantics as considered here. In any case, a semantics would help clarify our intuitions even about the pragmatic aspects of speech acts. As a clarification of my goals, note that the role of the proposed semantics is akin to that of classical semantics for assertives. Classical semantics only tells us when an assertive is objectively satisfied—it makes no claims about when an assertive should actually be uttered or believed.

Therefore, the focus here is on satisfaction conditions. The conditions of satisfaction for most kinds of speech acts differ significantly from those of assertives that are ordinarily considered in logic. Assertives, being claims of fact, are true or false; other speech acts call for a more complex framework in which their felicity or success can be described. In the context of imperatives, Hamblin distinguishes between what he calls *extensional* and *whole-hearted* satisfaction [Hamblin, 1987, pp. 153–157]. Briefly, the former notion admits accidental success, while the latter does not. Hamblin realized that these were useful things that may be said of a speech act; however, his aim was simply to be able state prescriptive conditions on when what kind of imperatives ought to be issued, and the philosophical problems that arise when one is in a "quandary." Thus his focus seems to have been pragmatic. I take advantage of some of his ideas, but make a finer distinction and extend it to other important kinds of speech acts here, formally relating them to intentions and know-how in the process. In §4.2, I address the problem of what kind of prescriptive constraints on communication may be stated, but see that as essentially supervenient on the semantics.

In §2, I discuss three different senses of satisfaction for the five kinds of speech acts considered in this paper. In §3, I describe a theory of intentions and know-how, and a formal model which I then use to formalize the different notions of satisfaction. In §4, I show how this theory may be used in the design of multiagent systems, and list some useful theorems.

2 Shades of Satisfaction

As remarked above, communication among agents in a multiagent system can be best understood by appealing to speech act theory [Austin, 1962; Searle, 1969]. In speech act theory an "illocution" (which I identify with a message) is seen to have two parts: an *illocutionary force* and a *proposition*. The illocutionary force distinguishes, e.g., a command from a promise; the proposition describes the state of the world that is, respectively, commanded or promised. This suggests a simple syntax for messages in an artificial language. A message, m, is a pair $\langle i, p\rangle$, where i identifies the illocutionary force, and p the proposition. Here i is an atomic symbol from the set {directive, commisive, permissive, prohibitive, assertive}; and p is a logical formula.

This much is quite standard even in the AI literature that deals with communication among agents [Huhns *et al.*, 1990; Thomas *et al.*, 1990]. However, none of the AI papers so far give a rigorous formal semantics for messages of different illocutionary forces. This lacuna is filled by this paper. But before I come to the formalization, I must discuss the different senses of *satisfaction* of speech acts. The rest of this section extends the discussion in [Hamblin, 1987, pp. 153–157]. Note that these different senses agree for the case of assertives.

The propositional part of a message specifies the state of the world that the message is, in some sense, about. An assertive asserts of that state that it holds (i.e., currently, though the proposition could be temporally indexed); a directive asks the hearer to bring that state about; a commisive commits the speaker to bringing it about; and so on. Thus the satisfaction of a message depends both on its illocutionary force and its proposition. The different notions of satisfaction are motivated using directives; other speech acts are considered in §3.

In the simplest sense of satisfaction, called *extensional* satisfaction in [Hamblin, 1987, p. 153], a message is said to be satisfied (with only minor qualifications) just if its proposition turns out to hold. E.g., a directive is satisfied when the proposition becomes true. Extensional satisfaction looks only at the immediate state of the world—the proposition may have been made true accidentally, or for the "wrong reasons," but would still meet the requirements for extensional satisfaction. This notion of satisfaction meets weak behavioral specifications; e.g., if the success of a speech act is part of a plan, then when it succeeds the agent can legitimately proceed to the next stage of the plan. However, this notion is not acceptable for complex systems because fortuitous circumstances would not, in general, arise often in them. A system whose agents were designed on this basis would turn out to be not sufficiently robust—e.g., we would expect a guarantee that some directives would be satisfied in a variety of circumstances, rather than that they once were. Also, it is of not much help practically, since it yields no insights about how the individual agents in a multiagent system ought to be designed.

This motivates the next sense of satisfaction, *whole-hearted* satisfaction. The whole-hearted satisfaction of a directive requires not only that the specified proposition be made true, but be made true in a sure-fire manner. The concerned agent should not only bring about the right state of the world, but know how to bring it about and intend to bring it about (thus it would bring it about in a way that exploits its know-how). By using whole-hearted satisfaction, the designer can require that an agent not issue two commands, which cannot both be whole-heartedly satisfied (due to limited know-how), even if they can both be extensionally satisfied.

Even whole-hearted satisfaction admits cases where the relevant proposition was made true, but only because it was going to be made true anyway, irrespective of whether the given speech act was performed. I.e., the speech act was pointless and played no real role in its satisfaction. This happens when, for a directive, the hearer was going to do the desired actions anyway. Often, it is useful to eliminate these conditions, so that the given speech act is really necessary. This requires that not only must the proposition in the speech act be made true in a sure-fire way, it must be made true *because* of that speech act. This, *relevant* satisfaction, is the strongest notion of satisfaction that I consider here.

The taxonomy of speech acts of this paper is motivated by the fact that permissives, prohibitives and directives have different satisfaction conditions (*cf.* [Bach and Harnish, 1979, pp. 39–54] and [Searle and Vanderveken, 1985, ch. 9], where permissives and prohibitives are lumped together with directives). The more convention- or culture-oriented illocutionary forces (e.g., christenings, greetings) are not considered here. Interrogatives are semantically quite like directives, but need special treatment to allow for answers; they are not included in this paper for reasons of space.

3 Formalization of Satisfaction Conditions

The different notions of satisfaction of speech acts depend on the definitions of know-how and intention. A formal rigorous theory of situated know-how and intentions has already been developed and reported. Here the same framework and technical definitions are used to give an account of the different sorts of satisfaction of several kinds of speech acts. This project has been inspired by that of C. L. Hamblin, who died while writing [Hamblin, 1987]. He argued that an account of imperatives must be built on top of a theory of abilities and intentions, especially, one which is "...not hidden in the mind, however, but expressed in action ..." [Hamblin, 1987, foreword by Belnap, p. viii]. He did not such have a formal theory of know-how and intentions, but I take this advice seriously, and try to give a semantics of several kinds of speech acts in terms of my earlier theory of intentions and know-how, where these concepts are defined in terms of the actions of agents situated in an objective model [Singh, 1990b; Singh, 1990a; Singh, 1991]. I briefly describe that theory, and then turn to the formal model.

3.1 Know-how and Intentions

This theory of know-how and intentions is meant to apply to both traditional plan-based architectures and modern situated ones. It uses the concept of *strategies* as abstract descriptions of the agents' behavior. Strategies correspond to plans in traditional systems and to the architectural structure of reactive agents, as instantiated at a given time. A strategy is simply the designer's description of the agent and the way in which it behaves. An agent knows how to achieve A, if it can achieve A whenever it so "intends." Let each agent have a strategy that it follows in the current situation. Intuitively, an agent knows how to achieve A relative to a strategy Y, iff it possesses the skills required to follow Y in such a way as to achieve A. Thus know-how is partitioned into two components: the "ability" to have satisfactory strategies, and the "ability" to follow them. Similarly, an agent intends to achieve A by performing a strategy Y, iff it currently has strategy Y and the successful performance of Y by it entails A; i.e., iff the agent is trying to perform Y to achieve A. Note that having an intention does not entail having the know-how to match it, and vice versa. Another important primitive is *can-prevent*, notated K_{prev}. This is related to know-how and applies when the given agent is able to perform actions so as to prevent the occurrence of the given condition. For reasons of space, the technical details of [Singh, 1991; Singh, 1990a; Singh, 1990b] are not included here. The presentation below is self-contained, however.

3.2 The Formal Model

The formal model here is based on possible worlds. Each possible *world* has a branching *history* of *times*. Agents influence the future by acting, but the outcome also depends on other events. A *scenario* at a given world and time is any linear branch of the future beginning there—this corresponds to a particular run or trace of the given system. A *subscenario* is a triple, $\langle S, t, t' \rangle$, which denotes a section of scenario S from time t to t'. The interpretation $[\![\,]\!]$ assigns sets of world-times pairs to predicates, for each possible tuple of their arguments; it assigns sets of subscenarios to each action, for each agent who might do it. Truth in the model, M, is defined relative to a world w and a time t: $M \models_{w,t} p$ denotes that p is true in M at w and t. Another useful notion is of truth relative to a scenario and a time in it (a scenario determines the world): $M \models_{S,t} p$ denotes that p is true in M on S at t. The formal language here is the predicate calculus, augmented with temporal logic (used in §4.3), and three predicates, 'K_{how},' 'intends' and 'K_{prev},' each applying to an agent, a strategy and a formula.

Speech acts are, first of all, actions. For simplicity, they are seen as the actions of just their speakers, and occur over subscenarios. Let 'says-to' be a parametrized speech act, to be used as 'says-to(y, m).' This action will be seen as an action done by agent x. $\langle t_s, t_h \rangle \in [\![\text{says-to}(y, m)]\!]^x$ means that agent x communicated message m to agent y in the time from t_s (the time of speaking) to t_h (the time of hearing). This just says that the message was successfully transmitted (it is possible to allow failed transmissions, but that is not useful here). Define a new predicate 'comm' that applies to two agents, and a message. 'Comm(x, y, m)' is true at w, t just if x said (or started to say) m to y then. A transmitted message may, of course, not be satisfiable. In order to be able to talk of the different kinds of satisfaction of messages, I introduce three operators: ESAT, WSAT and RSAT (collectively called SAT below) that apply on formulas of the form 'comm(x, y, m).' They respectively state that the given message is extensionally, whole-heartedly, and relevantly satisfied. Now the conditions of truth are given for each of these three operators and for each of the possible illocutionary forces, relative to a scenario and a time.

The definition of truth is standard for the classical and temporal parts of the logic, and is not given here to save space. For p not of the form $\mathsf{SAT}q$, $M \models_{S,t} p$ iff $M \models_{w,t} p$, where S is a scenario at w, t. For p of the form $\mathsf{SAT}q$, it is convenient to give these definitions relative to a scenario and time, rather than directly relative to a world and time. We would like to define $M \models_{w,t} p$ as $M \models_{S,t} p$, where S is restricted in some way. S has to be restricted because there is always (in practice) a scenario where $\mathsf{SAT}q$ would fail. One natural restriction is to scenarios compatible with the speaker's current intentions (or strategy). This ensures $\mathsf{SAT}q$ as long as the speaker's intentions do not change.

3.3 Extensional Satisfaction

Extensional satisfaction is defined relative to a scenario and a time in it (thus "future" means future within that scenario). A directive is satisfied at a scenario and time just if its proposition becomes true at a future time on that scenario. A commisive too is satisfied just if its proposition becomes true at some future time on the given scenario. An assertive is satisfied just at those times where its proposition is true. A permissive is satisfied on a scenario and time just if it is taken advantage of sometimes in the future of that time on that scenario. A prohibitive is satisfied just if it is never violated in the future of the given time on the given scenario. Thus the agents' intentions and know-how do not matter for extensional satisfaction.

1. $M \models_{S,t} \mathsf{ESAT}(\text{comm}(x, y, \langle\text{directive}, p\rangle))$ iff $(\exists t' \in S : t' \geq t \wedge M \models_{S,t'} p)$
2. $M \models_{S,t} \mathsf{ESAT}(\text{comm}(x, y, \langle\text{commisive}, p\rangle))$ iff $(\exists t' \in S : t' \geq t \wedge M \models_{S,t'} p)$
3. $M \models_{S,t} \mathsf{ESAT}(\text{comm}(x, y, \langle\text{permissive}, p\rangle))$ iff $(\exists t' \in S : t' \geq t \wedge M \models_{S,t'} p)$
4. $M \models_{S,t} \mathsf{ESAT}(\text{comm}(x, y, \langle\text{prohibitive}, p\rangle))$ iff $(\forall t' \in S : t' > t \rightarrow M \not\models_{S,t'} p)$
5. $M \models_{S,t} \mathsf{ESAT}(\text{comm}(x, y, \langle\text{assertive}, p\rangle))$ iff $M \models_{S,t} p$

3.4 Whole-hearted Satisfaction

Whole-hearted satisfaction too is defined relative to a scenario and a time. A directive is satisfied on a scenario and time just if its proposition becomes true at a future time in that scenario, and all along the scenario from the given time to then, the hearer has the know-how, as well as the intention to achieve it. I.e., if the hearer has a strategy (that it may be said to be following) relative to which it has the know-how and the intention to achieve p (as explained in §3.1). Similarly, a commisive is satisfied just when its proposition becomes true at some future time on the given scenario, and all along the scenario from the given time to then, the *speaker* has the know-how to achieve it and also intends it. The condition for assertives is unchanged.

A permissive is satisfied at a scenario and a time just if it is taken advantage of by the hearer at a future point on that scenario. But when a permissive is taken advantage of, it allows the hearer to do actions at certain times that it could not have done before because they might possibly lead to the condition becoming true. Thus a permissive is satisfied on a scenario on which the hearer does at least one action whose performance can lead to a situation where it is unable to prevent that condition from occurring (i.e., the hearer can now risk letting that condition hold). Similarly, a prohibitive is satisfied at a scenario and time just if none of the actions done by the hearer on that scenario (in the future), can lead to a situation where the hearer would be unable to prevent the condition from occurring (i.e., the hearer does not risk violating the prohibition).

1. $M \models_{S,t} \mathsf{WSAT}(\text{comm}(x, y, \langle\text{directive}, p\rangle))$ iff $(\exists t' \in S : t' \geq t \wedge M \models_{S,t'} p \wedge (\forall t'' : t \leq t'' < t' \rightarrow (\exists Y : M \models_{S,t''} K_{how}(y, Y, p) \wedge \text{intends}(y, Y, p))))$
2. $M \models_{S,t} \mathsf{WSAT}(\text{comm}(x, y, \langle\text{commisive}, p\rangle))$ iff $(\exists t' \in S : t' \geq t \wedge M \models_{S,t'} p \wedge (\forall t'' : t \leq t'' < t' \rightarrow (\exists Y : M \models_{S,t''} K_{how}(x, Y, p) \wedge \text{intends}(x, Y, p))))$
3. $M \models_{S,t} \mathsf{WSAT}(\text{comm}(x, y, \langle\text{permissive}, p\rangle))$ iff $(\exists t' \in S : t' \geq t \wedge (\forall a : (\exists t'' : \langle S, t', t''\rangle \in [\![a]\!]^y) \rightarrow (\exists S', t''' : t''' \in S' \wedge \langle S', t', t'''\rangle \in [\![a]\!]^y \wedge (\forall Y : M \not\models_{S',t'''} K_{prev}(y, Y, p)))))$
4. $M \models_{S,t} \mathsf{WSAT}(\text{comm}(x, y, \langle\text{prohibitive}, p\rangle))$ iff $(\forall t' \in S : t' > t \rightarrow (\forall a : (\exists t'' : \langle S, t', t''\rangle \in [\![a]\!]^y) \rightarrow (\forall S', t''' : t''' \in S' \wedge \langle S', t', t'''\rangle \in [\![a]\!]^y \rightarrow (\exists Y : M \models_{S',t'''} K_{prev}(y, Y, p)))))$

3.5 Relevant Satisfaction

Relevant satisfaction is also defined relative to a scenario and a time. It resembles the previous case, but differs in adding a requirement that roughly says that the given speech act is the true reason for its success. A directive is satisfied just when its proposition becomes true at a time in the future of the given time, and all along the scenario from the given time to then, the hearer has the know-how to achieve it, and furthermore that this know-how does not arise in at least one scenario that is a temporal alternative to the given one. The definition for commisives is analogous. The condition for assertives continues to be the same as before.

A permissive is satisfied at a scenario and a time just if it is taken advantage of by the hearer at a future time on the given scenario, with an additional requirement that that there be an alternative scenario to the given one where it is not taken advantage of. As before, when a permissive is taken advantage of, it is WSAT. However, in this case, it must also be the case that on at least one other scenario, the permissive is not WSAT. Intuitively, this scenario is one of those where a stronger permission is not in force. The condition for prohibitives parallels that for permissives. Let $\mathbf{S}_{w,t}$ be the set of all scenarios beginning at w, t; below, let $S \in \mathbf{S}_{w,t}$.

1. $M \models_{S,t} \mathsf{RSAT}(\mathrm{comm}(x,y,m))$ iff $M \models_{S,t} \mathsf{WSAT}(\mathrm{comm}(x,y,m)) \wedge (\exists S' \in \mathbf{S}_{w,t} : M \not\models_{S',t} \mathsf{WSAT}(\mathrm{comm}(x,y,m)))$

This definition could have been strengthened by adding a notion of the "closeness" of scenarios to each other, so that we could say that S' above was not just any scenario but one of the ones closest to S. However, this idea is not pursued in this paper.

4 Applying the Theory

The two main motivations for developing this theory are to provide a rigorous foundation for the design of multiagent systems and to justify some prescriptive claims about how agents should communicate in them. This theory has given objective criteria to evaluate the correctness of different scenarios, or runs of a multiagent system. In design, the problem is to create a system which allows only correct scenarios to be actualized. Prescriptive claims for agents tell them what to do given their beliefs and intentions, so that correct scenarios emerge. I discuss these two problems below, and then some important formal consequences of this theory.

4.1 Designing Multiagent Systems

One extension needed is to distinguish between messages of the same major illocutionary force that are different in some important respects. E.g., commands differ from requests, since they presuppose authority on part of the speaker, and can cause a change in the hearer's intentions under a wider variety of circumstances than requests. A correct scenario with a command must WSAT it as a directive (assuming authority); one with a request must WSAT it too, but only if some other conditions on cooperation are met. The designer needs to constrain the issuing of directives, and/or increase know-how and constrain intentions so that the system actualizes only correct scenarios. E.g., requiring that agents persist with their strategies for sufficiently long considerably simplifies generating runs that on which directives and commisives are WSAT. Correctness must be ensured for all the message types that can occur in the system.

As an example, consider the contract net [Davis and Smith, 1983]. In its simplest form, a *manager* sends out a call for bids to all *contractors*—treat this as requesting the hearer to bid and to promise that it will do the task for a certain price (if it wants to bid). Contractors who bid thus promise that if requested to do the task and given their price, they will do it. The manager selects one contractor and requests it to do the task. In any correct design for this protocol, the contractors must have (at the appropriate times) (1) the know-how to answer correctly whether they will bid on a given task, and to promise as described above; and (2) the corresponding intentions. Thus the design goal reduces to ensuring conditions (1) and (2) for all the agents.

The different senses of SAT yield rigorous definitions for three kinds of correctness conditions for multiagent systems. A scenario can be defined to be correct in the sense of SAT if all the messages passed on it are satisfied in the same sense. In general terms, the designer's goal is to ensure that all runs that may be actualized are correct. This reduces to the goal that the intentions and know-hows of the agents are such that only correct scenarios are actualized. This is the sense of correctness that designers use in practice. They usually achieve this kind of correctness by, e.g., hard-wiring the intention to cooperate in their agents, or by setting up hierarchical structures such that some directives (commands) are always obeyed, and others (requests) obeyed whenever they do not conflict with the hearer's current intentions.

4.2 Normative Constraints on Communication

The objective criteria given above can be used to motivate some normative constraints on communication among the agents in a multiagent system. These constraints could be used by the designer, and possibly by the agents themselves to reason about the messages they exchange. Note that these are just meant as weak constraints, and may be easily overridden. If (some of) these constraints are obeyed, the scenarios that are actualized are not just correct but also "good." Imposing these constraints can simplify a design since certain good properties can then be taken for granted.

1. An agent should issue a directive only if its intentions are satisfied (i.e., its current strategies are performed successfully) on all scenarios on which the directive is WSAT. This is the converse of the definition for $M \models_{w,t} \mathsf{WSAT}q$ suggested in §3.2—this ensures that actions by the hearer will not render the speaker's own intentions impossible to achieve.
2. All messages sent must be RSAT, where the quantification over scenarios is restricted to scenarios compatible with the speaker's current intentions.
3. Agents ought to persist with their strategies (i.e., their strategy at a later time should be the appropriate "tail" of their strategy at an earlier time). This not only simplifies the WSAT and RSAT conditions for directives and commisives, but also simplifies the interactions among agents by, e.g., making it easy for an issuer of a directive to not take on a strategy that would interfere with its compliance.
4. All messages sent by one speaker must be mutually consistent in the sense of being jointly satisfiable on at least some scenarios. E.g., different directives should not clash with each other and prohibitives should not preclude the satisfaction of directives and other prohibitives. This prevents many unacceptable situations, but can cause problems if some redundant permissions are issued, i.e., those that might never be used on some scenarios where the directives are met. So one should exclude (such) permissions from this constraint.

4.3 Some Theorems

A useful feature of the present theory is that it brings the different kinds of SAT within the fold of logic. In order to write some formal consequences of the preceding definitions, I will need to define some temporal operators. Here E means "in *some* scenario beginning at this time"; A $\equiv \neg$E$\neg$; pUq "q holds in the future and p holds until then"; P "at a past time up to the beginning of this scenario"; F$p \equiv$ **true**Up "eventually p"; and G$p \equiv \neg$F$\neg p$ "always p" [Emerson, 1990]. E evaluated at a scenario is evaluated at the first point in that scenario.

1. RSAT entails WSAT; WSAT entails ESAT.
2. For a given scenario, ESAT(comm($x, y, \langle i, p\rangle$)) $\equiv$ Fp for $i \neq$ prohibitive. For prohibitives, it $\equiv$ G$\neg p$.
3. RSAT(comm($x, y, \langle i, p\rangle$)) $\equiv$ WSAT(comm($x, y, \langle i, p\rangle$)) $\wedge$ E$\neg$WSAT(comm($x, y, \langle i, p\rangle$))
   This follows easily from the definition.
4. WSAT(comm(x, y, ⟨directive,p⟩)) $\equiv$ ESAT(comm(x, y, ⟨directive,p⟩)) $\wedge$ ($\exists Y : (K_{how}(y, Y, p) \wedge$ intends(y, Y, p)))Up
   A directive is WSAT iff it is ESAT and until it is, the hearer intends to and knows how to achieve it.
5. WSAT(comm(x, y, ⟨directive,p⟩)) $\equiv$ ESAT(comm(x, y, ⟨directive,p⟩)) $\wedge$ ESAT(comm(x, y, ⟨prohibitive, $\neg(\exists Y : K_{how}(y, Y, p) \wedge$ intends(y, Y, p)) $\wedge \neg$Pp⟩))
   This is an important intuitive result since it shows that a directive interpreted in the WSAT sense has the force of a directive in the ESAT sense conjoined with a *prohibitive* in the ESAT sense.

5 Conclusions

None of the extant theories of communication address quite the same problem as I have addressed in this paper. The theory presented here refines and formalizes some intuitions about communication among agents. I started out with a set of obvious and well-known intuitions about the nature of communication derived from classical speech act theory. However, I motivated and approached an important problem that has not been addressed in the literature on speech act theory, or even in the AI literature. My commitment is stronger to the general claims than to the specifics. However, using definitions of the intentions and know-how of an agent, I was systematically able to give rigorous definitions of the conditions of satisfaction for speech acts of different of illocutionary forces. These definitions capture many of our intuitions about when, as speakers and hearers, we believe that a given speech act has been satisfied. This theory has applications in the design of multiagent systems, where constraints on the know-how and intentions of agents are derived from their desired communicative behavior. It can also yield some well-motivated normative constraints on communication among agents. An advantage of the formal approach is that this process can be guided by the several theorems that exist.

References

[Allen and Perrault, 1980] James F. Allen and C. Raymond Perrault. Analyzing intention in utterances. *Artificial Intelligence*, 15:143–178, 1980.

[Austin, 1962] John L. Austin. *How to do Things with Words*. Clarendon, Oxford, UK, 1962.

[Bach and Harnish, 1979] Kent Bach and Robert M. Harnish. *Linguistic Communication and Speech Acts*. MIT Press, Cambridge, MA, 1979.

[Cohen and Levesque, 1988] Philip R. Cohen and Hector J. Levesque. Rational interaction as the basis for communication. Technical Report 433, SRI International, Menlo Park, CA, April 1988.

[Davis and Smith, 1983] Randall Davis and Reid G. Smith. Negotiation as a metaphor for distributed problem solving. *Artificial Intelligence*, 20:63–109, 1983. Reprinted in *Readings in Distributed Artificial Intelligence*, A. H. Bond and L. Gasser, eds., Morgan Kaufmann, 1988.

[Emerson, 1990] E. A. Emerson. Temporal and modal logic. In J. van Leeuwen, editor, *Handbook of Theoretical Computer Science*, volume B. North-Holland Publishing Company, Amsterdam, The Netherlands, 1990.

[Gasser and Huhns, 1989] Les Gasser and Michael N. Huhns, editors. *Distributed Artificial Intelligence, Volume II*. Pitman/Morgan Kaufmann, London, 1989.

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

[Hamblin, 1987] C. L. Hamblin. *Imperatives*. Basil Blackwell Ltd., Oxford, UK, 1987.

[Huhns *et al.*, 1990] Michael N. Huhns, David Bridgeland, and Natraj Arni. A DAI communication aide. Technical Report ACT-RA-317-90, Microelectronics and Computer Technology Corporation, Austin, TX, October 1990.

[Huhns, 1987] Michael N. Huhns, editor. *Distributed Artificial Intelligence*. Pitman/Morgan Kaufmann, London, 1987.

[Perrault, 1987] Raymond Perrault. An application of default logic to speech act theory. Technical Report 90, Center for the Study of Language and Information, Stanford, CA, March 1987.

[Searle and Vanderveken, 1985] John R. Searle and Daniel Vanderveken. *Foundations of Illocutionary Logic*. Cambridge University Press, Cambridge, UK, 1985.

[Searle, 1969] John R. Searle. *Speech Acts*. Cambridge University Press, Cambridge, UK, 1969.

[Searle, 1975] John R. Searle. Indirect speech acts. In P. Cole and J. L. Morgan, editors, *Syntax and Semantics, Volume 3*. Academic Press, New York, NY, 1975. Reprinted in [?].

[Singh, 1990a] Munindar P. Singh. Group intentions. In *10th Workshop on Distributed Artificial Intelligence*, October 1990.

[Singh, 1990b] Munindar P. Singh. Towards a theory of situated know-how. In *9th European Conference on Artificial Intelligence*, August 1990.

[Singh, 1991] Munindar P. Singh. A logic of situated know-how. In *National Conference on Artificial Intelligence (AAAI)*, July 1991.

[Thomas *et al.*, 1990] Becky Thomas, Yoav Shoham, and Anton Schwartz. Modalities in agent-oriented programming. Computer Science Department, Stanford University, February 1990.