A First-Order Theory of Communicating First-Order Formulas

Ernest Davis*
Courant Institute
New York University
davise@cs.nyu.edu

Abstract

This paper presents a theory of informative communications among agents that allows a speaker to communicate to a hearer truths about the state of the world; the occurrence of events, including other communicative acts; and the knowledge states of any agent — speaker, hearer, or third parties — any of these in the past, present, or future — and any logical combination of these, including formulas with quantifiers. We prove that this theory is consistent, and compatible with a wide range of physical theories. We examine how the theory avoids two potential paradoxes, and discuss how these paradoxes may pose a danger when this theory are extended.

Keywords: Communication, knowledge, paradox.

Introduction

In constructing a formal theory of communications between agents, the issue of expressivity enters at two different levels: the scope of what can be said *about* the communications, and the scope of what can be said *in* the communications. Other things being equal, it is obviously desirable to make both of these as extensive as possible. Ideally, a theory should allow a speaker to communicate to a hearer truths about the state of the world; the occurrence of events, including other communicative acts; the knowledge states of any agent — speaker, hearer, or third parties; any of these in the past, present, or future; and any logical combination of these. This paper presents a theory that achieves pretty much that.

A few examples of what can be expressed, together with their formal representation:

1. Alice tells Bob that all her children are asleep.

$$\begin{aligned}&\exists_Q \text{ occurs(do(alice,inform(bob,}Q\text{)),s0,s1)} \wedge \\ &\quad \forall_S \text{ holds}(S,Q) \Leftrightarrow \\ &\quad [\forall_C \text{ holds}(S,\text{child}(C,\text{alice})) \Rightarrow \\ &\qquad \text{holds}(S,\text{asleep}(C))].\end{aligned}$$

2. Alice tells Bob that she doesn't know whether he locked the door.

$$\begin{aligned}&\exists_Q \text{ occurs(do(alice,inform(bob,}Q\text{)),s0,s1)} \wedge \\ &\quad \forall_S \text{ holds}(S,Q) \Leftrightarrow \\ &\qquad [\exists_{SA} \text{ k\_acc(alice},S,SA) \wedge \\ &\qquad\quad \exists_{S1A,S2A}\ S1A < S2A < SA \wedge \\ &\qquad\qquad \text{occurs(do(bob,lock\_door)},S1A,S2A)] \wedge \\ &\qquad [\exists_{SA} \text{ k\_acc(alice},S,SA) \wedge \\ &\qquad\quad \neg\exists_{S1A,S2A}\ S1A < S2A < SA \wedge \\ &\qquad\qquad \text{occurs(do(bob,lock\_door)},S1A,S2A)].\end{aligned}$$

3. Alice tells Bob that if he finds out who was in the kitchen at midnight, then he will know who killed Colonel Mustard. (Note: The interpretation below assumes that exactly one person was in the kitchen at midnight.)

$$\begin{aligned}&\exists_Q \text{ occurs(do(alice,inform(bob,}Q\text{)),s0,s1)} \wedge \\ &\quad \forall_S \text{ holds}(S,Q) \Leftrightarrow \\ &\qquad \forall_{S2}\ [S2 > S \wedge \\ &\qquad\quad \exists_{PK} \forall_{S2A} \text{ k\_acc(bob},S2,S2A) \Rightarrow \\ &\qquad\qquad \exists_{S3A}\ S3A < S2A \wedge \\ &\qquad\qquad\quad \text{midnight(time}(S3A)) \wedge \\ &\qquad\qquad\quad \text{holds}(S3A,\text{in}(PK,\text{kitchen}))] \Rightarrow \\ &\qquad\quad [\exists_{PM} \forall_{S2B} \text{ k\_acc(bob},S2,S2B) \Rightarrow \\ &\qquad\qquad \exists_{S3B,S4B}\ S3B < S4B < S2B \wedge \\ &\qquad\quad \text{occurs(do}(PM,\text{murder(mustard))},S3B,S4B)].\end{aligned}$$

4. Alice tells Bob that no one had ever told her she had a sister.

$$\begin{aligned}&\exists_Q \text{ occurs(do(alice,inform(bob,}Q\text{)),s0,s1)} \wedge \\ &\quad \forall_S \text{ holds}(S,Q) \Leftrightarrow \\ &\qquad \neg\exists_{S2,S3,Q1,P1}\ S2 < S3 < S \wedge \\ &\qquad\ \text{occurs(do}(P1,\text{inform(alice},Q1)),S2,S3) \wedge \\ &\qquad\ \forall_{SX} \text{ holds}(SX,Q1) \Rightarrow \\ &\qquad\qquad \exists_{P2} \text{ holds}(SX,\text{sister}(P2,\text{alice})).\end{aligned}$$

5. Alice tells Bob that he has never told her anything she didn't already know.

$$\begin{aligned}&\exists_Q \text{ occurs(do(alice,inform(bob,}Q\text{)),s0,s1)} \wedge \\ &\quad \forall_S \text{ holds}(S,Q) \Leftrightarrow \\ &\qquad \forall_{S2,S3,Q1} \\ &\qquad\quad [S2 < S3 \leq S \wedge \\ &\qquad\quad \text{occurs(do(bob,inform(alice},Q1)),S2,S3)] \Rightarrow \\ &\qquad\quad \forall_{S2A} \text{ k\_acc(alice},S2,S2A) \Rightarrow \text{holds}(S2A,Q1).\end{aligned}$$

*The research reported in this paper was supported in part by NSF grant IIS-0097537. The work described here comes out of and builds upon a project done in collaboration with Leora Morgenstern, stemming from a benchmark problem that she proposed. Thanks also to Carl Woolf, who first showed me the unexpected hanging problem many years ago.

These representations works as follows: The expression "do(AS,inform(AH, Q))" denotes the action of speaker AS informing AH that fluent Q holds in the current situation. The content Q here is a *generalized fluent*, that is, a property of situations / possible worlds. Simple fluents are defined by ground terms, such as "in(mustard,kitchen)." In more complex cases, the fluent Q is characterized by a formula "$\forall_S$ holds(S, Q) $\Leftrightarrow \alpha(S)$" where α is some formula open in S. (Equivalently, Q could be defined using the lambda expression Q=$\lambda(S)\alpha(S)$.)

The above examples illustrate many of the expressive features of our representation:

- Example 1 shows that the content of a communication may be a quantified formula.
- Example 2 shows that the content of a communication may refer to knowledge and ignorance of past actions.
- Example 3 shows that the content of a communication may be a complex formula involving both past and present events and states of knowledge.
- Examples 4 and 5 show that the content of a communication may refer to other communications. They also show that the language supports quantification over the content of a communication, and thus allows the content to be partially characterized, rather than fully specified.

If we wish to reason about such informative actions — e.g. to be sure that they can be executed — then we must be sure, among other conditions, that the fluent denoting the content of the action exists. This requires a comprehension axiom that asserts that such a fluent exists for *any* such formula α. Comprehension axioms often run the risk of running into analogues of Russell's paradox, but this one turns out to be safe. We will discuss two paradoxes that look dangerous for this theory, but the theory succeeds in side-stepping these. One of these is the well-known "unexpected hanging" paradox. To make sure that there are no further paradoxes in hiding that might be more destructive, we prove that our theory is consistent, and compatible with a wide range of physical theories.

The paper proceeds as follows: We first discuss the theories of time, of knowledge, and of communication that we use. We illustrate the power of the theory by showing how it supports two example inferences. We describe an apparent paradox and how it is avoided. We show how the theory avoids the "unexpected hanging" problem. We present the proof that the theory is consistent. We discuss related work and future work and present our conclusions.

Framework

We use a situation-based, branching theory of time; an interval-based theory of multi-agent actions; and a possible-worlds theory of knowledge. This is all well known, so the description below is brief.

Time and Action

We use a situation-based theory of time. Time can be either continuous or discrete, but it must be *branching*, like the situation calculus. The branching structure is described by the

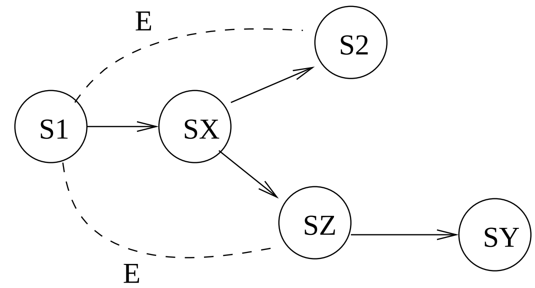


Figure 1: Axiom T.9

partial ordering "$S1 < S2$", meaning that there is a timeline containing $S1$ and $S2$ and $S1$ precedes $S2$. It is convenient to use the abbreviations "$S1 \leq S2$" and "ordered($S1, S2$)." The predicate "holds(S, Q)" means that fluent Q holds in situation S.

Each agent has, in various situations, a choice about what action to perform next, and the time structure includes a separate branch for each such choice. Thus, the statement that action E is possible in situation S is expressed by asserting that E occurs from S to $S1$ for some $S1 > S$.

Following (McDermott 1982), actions are represented as occurring over an interval; the predicate occurs($E, S1, S2$) states that action E occurs starting in $S1$ and ending in $S2$. However, the whole theory could be recast without substantial change into the situation calculus extended to permit multiple agents, after the style of (Reiter, 2001).

Table 1 shows the axioms of our temporal theory. Throughout this paper, we use a sorted first-order logic with equality, where the sorts of variables are indicated by their first letter. The sorts are clock-times (T), situations (S), Boolean fluents (Q), actions (E), agents (A), and actionals (Z). (The examples at the beginning of this paper use some terms of other sorts *ad hoc*; these are self-explanatory.) An *actional* is a characterization of an action without specifying the agent. For example, the term "puton(blocka,table)" denotes the actional of someone putting block A on the table. The term "do(john, puton(blocka,table))" denotes the action of John putting block A on the table. Free variables in a formula are assumed to be universally quantified.

Our theory does not include a representation of what *will* happen from a given situation as opposed to what *can* happen. This will be important in our discussion of the paradoxes.

Knowledge

As first proposed by Moore (1980,1985) and widely used since, knowledge is represented by identifying temporal situations with epistemic possible worlds and positing a relation of knowledge accessibility between situations. The relation k_acc(A, S, SA) means that situation SA is accessible from S relative to agent A's knowledge in S; that is, as far as A knows in S, the actual situation could be SA. The statement that A knows ϕ in S is represented by asserting that ϕ holds in every situation that is knowledge accessible from S for A. As is well known, this theory enables the expression

Primitives:

$T1 < T2$ — Time $T1$ is earlier than $T2$.

$S1 < S2$ — Situation $S1$ precedes $S2$, on the same time line. (We overload the $<$ symbol.)

time(S) — Function from a situation to its clock time.

holds(S, Q) — Fluent Q holds in situation S.

occurs($E, S1, S2$) — Action E occurs from situation $S1$ to situation $S2$.

do(A, Z) — Function. The action of agent A doing actional Z.

Definitions:

TD.1 $S1 \leq S2 \equiv S1 < S2 \vee S1 = S2$.

TD.2 ordered($S1, S2$) $\equiv$
$S1 < S2 \vee S1 = S2 \vee S2 < S1$.

Axioms:

T.1 $T1 < T2 \vee T2 < T1 \vee T1 = T2$.

T.2 $\neg[T1 < T2 \wedge T2 < T1]$.

T.3 $T1 < T2 \wedge T2 < T3 \Rightarrow T1 < T3$.
(Clock times are linearly ordered)

T.4 $S1 < S2 \wedge S2 < S3 \Rightarrow S1 < S3$. (Transitivity)

T.5 $(S1 < S \wedge S2 < S) \Rightarrow$ ordered$(S1, S2)$.
(Forward branching)

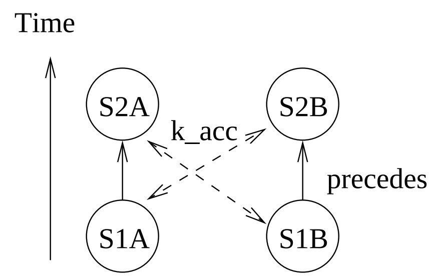
T.6 $S1 < S2 \Rightarrow$ time($S1$) $<$ time($S2$).
(The ordering on situations is consistent with the orderings of their clock times.)

T.7 $\forall_{S,T1}\, \exists_{S1}$ ordered($S, S1$) $\wedge$ time($S1$)=$T1$.
(Every time line contains a situation for every clock time.)

T.8 occurs($E, S1, S2$) $\Rightarrow S1 < S2$.
(Events occur forward in time.)

T.9 [occurs($E, S1, S2$) $\wedge\, S1 < SX < S2 \,\wedge$
$SX < SY] \Rightarrow$
$\exists_{SZ}$ ordered(SY, SZ) $\wedge$ occurs($E, S1, SZ$).
(If action E starts to occur on the time line that includes SY, then it completes on that time line. (Figure 1))

Table 1: Temporal Axioms



Axiom K.6 prohibits this structure.

Figure 2: Axiom K.6

of complex interactions of knowledge and time; one can represent both knowledge about change over time and change of knowledge over time.

Again following Moore (1985), the state of agent A knowing *what something is* is expressed by using a quantifier of larger scope than the universal quantification over accessible possible worlds. For example, the statement, "In situation s1, John knows who the President is" is expressed by asserting that there exists a unique individual who is the President in all possible worlds accessible for John from s1.

$\exists_X\, \forall_{S1A}$ k_acc(john,s1,$S1A$) $\Rightarrow$
holds($S1A$,president(X)).

For convenience, we posit an S5 logic of knowledge; that is, the knowledge accessibility relation, restricted to a single agent, is in fact an equivalence relation on situations. This is expressed in axioms K.1, K.2, and K.3 in table 2. Three important further axioms govern the relation of time and knowledge.

K.4. Axiom of memory: If A knows ϕ in S, then in any later situation, he remembers that he knew ϕ in S.

K.5. A knows all the actions that he has begun, both those that he has completed and those that are ongoing. That is, he knows a *standard identifier* for these actions; if Bob is dialing (212) 998-3123 on the phone, he knows that he is dialing (212) 998-3123 but he may not know that he is calling Ernie Davis. At any time, A knows what actions he can now begin.

K.6 Knowledge accessibility relations do not cross in the time structure. I have not found any natural expression of this axiom, but certainly a structure that violated it would be a very odd one. (Figure 2.)

The theory includes a forms of common knowledge, restricted to two agents. Agents $A1$ and $A2$ have *shared knowledge* of ϕ if they both know ϕ, they both know that they both know ϕ and so on. We represent this by defining a further accessibility relation, "sk_acc($A1, A2, S, SA$)" (SA is accessible from S relative to the shared knowledge of $A1$ and $A2$). This is defined as the transitive closure of links of the form k_acc($A1, \cdot, \cdot$) together with links of the form k_acc($A2, \cdot, \cdot$). (Of course, transitive closure cannot be exactly defined in a first-order theory; we define an approximation that is adequate for our purposes.)

Primitives:
k_acc(A, SA, SB) — SB is accessible from SA relative to A's knowledge in SA.
sk_acc($A1, A2, SA, SB$) — SB is accessible from SA relative to the shared knowledge of $A1$ and $A2$ in SA.

Axioms

K.1 $\forall_{A,SA}$ k_acc(A, SA, SA).

K.2 k_acc(A, SA, SB) $\Rightarrow$ k_acc(A, SB, SA)

K.3 k_acc(A, SA, SB) $\wedge$ k_acc(A, SB, SC) $\Rightarrow$ k_acc(A, SA, SC).
(K.1 through K.3 suffice to ensure that the knowledge of each agent obeys an S5 logic: what he knows is true, if he knows ϕ he knows that he knows it; if he doesn't know ϕ, he knows that he doesn't know it.)

K.4 [k_acc($A, S2A, S2B$) $\wedge$ $S1A < S2A$] $\Rightarrow$
$\exists_{S1B}$ $S1B < S2B$ $\wedge$ k_acc($A, S1A, S1B$).
(Axiom of memory: If agent A knows ϕ at any time, then at any later time he knows that ϕ was true.)

K.5 [occurs(do(A, Z),$S1A, S2A$) $\wedge$ $S1A \leq SA$ $\wedge$
ordered($SA, S2A$) $\wedge$ k_acc(A, SA, SB)] $\Rightarrow$
$\exists_{S1B,S2B}$ occurs(do(A, Z),$S1B, S2B$) $\wedge$
$S1B \leq SB$ $\wedge$
$[S2A < SA \Rightarrow S2B < SB]$ $\wedge$
$[S2A = SA \Rightarrow S2B = SB]$ $\wedge$
$[SA < S2A \Rightarrow SB < S2B]$ $\wedge$
$[S1A = SA \Rightarrow S1B = SB]$
(An agent knows all the actions that he has begun, and all the actions that are feasible now, and the state of their completion.)

K.6 $\neg\exists_{A,S1A,S1B,S2A,S2B}$ $S1A < S2A \wedge S1B < S2B$ $\wedge$ k_acc($A, S1A, S2B$) $\wedge$ k_acc($A, S2A, S1B$).
(Knowledge accessibility links do not cross in the time structure (Figure 2.)

K.7 sk_acc($A1, A2, SA, SB$) $\Leftrightarrow$
[k_acc($A1, SA, SB$) $\vee$ k_acc($A2, SA, SB$) $\vee$
sk_acc($A1, A2, SB, SA$) $\vee$
sk_acc($A2, A1, SA, AB$) $\vee$
$\exists_{SC}$ sk_acc($A1, A2, SA, SC$) $\wedge$
sk_acc($A1, A2, SC, SB$)].
Definition of sk_acc as a equivalence relation, symmetric in $A1, A2$, that includes the k_acc links for the two agents $A1, A2$.

Table 2: Axioms of Knowledge

Communication

We now introduce the function "inform", taking two arguments, a agent AH and a fluent Q. The term "inform(AH, Q)" denotes the actional of informing AH that Q; the term "do(AS,inform(AH, Q))" thus denotes the action of speaker AS informing AH that Q. Our theory here treats "do(AS,inform(AH, Q))" as a primitive actions; in a richer theory, it would be viewed as an illocutionary description of an underlying locutionary act (not here represented) — the utterance or writing or broadcasting of a physical signal.

We also add a second actional "communicate(AH)". This alternative characterization of a communicative act, which specifies the hearer but not the content of the communication, enables us to separate out *physical* constraints on a communicative act from *contentive* constraints. Thus, we allow a purely physical theory to put constraints on the occurrence of a communication, or even to posit physical effects of a communication, but these must be independent of the information content of the communication.

We posit the following axioms:

I.1 Any inform act is a communication.
occurs(do(AS,inform(AH, Q)),$S1, S2$) $\Rightarrow$
occurs(do(AS,communicate(AH)),$S1, S2$).

I.2. If a speaker AS can communicate with a hearer AH, then AS can inform AH of some specific Q if and only if A knows that Q holds at the time he begins speaking.

[$\exists_{SX}$ occurs(do(AS,communicate(AH)),$S1, SX$)] $\Rightarrow$
[$\forall_Q$ [$\exists_{S2}$ occurs(do(AS,inform(AH, Q)),$S1, S2$)] $\Leftrightarrow$
[$\forall_{S1A}$ k_acc($AS, S1, S1A$) $\Rightarrow$ holds($S1A, Q$)]]

I.3. If AS informs AH of Q from $S1$ to $S2$, then in $S2$, AH and AS have shared knowledge that this event has occurred. It follows from I.3, I.2, and K.5 that in $S2$, AS and AH have shared knowledge that Q held in $S1$. (See Lemma 1, below).

$\forall_{S1,S2,S2A}$ [occurs(do(AS,inform(AH, Q)),$S1, S2$) $\wedge$
sk_acc($AS, AH, S2, S2A$)] $\Rightarrow$
$\exists_{S1A}$ occurs(do(AS,inform(AH, Q)),$S1A, S2A$).

(If axiom K.7 were replaced by a second-order axiom stating that sk_acc was the true transitive closure of k_acc, then it would suffice here to say that AH knows that the inform act has occurred.)

I.4. If AS informs AH of $Q1$ over $[S1, S2]$ and the shared knowledge of AS and AH in $S1$ implies that holds($S1, Q1$) $\Leftrightarrow$ holds($S1, Q2$), then AS has also informed AH of $Q2$ over $[S1, S2]$. Conversely, the two actions "do(AS,inform($AH, Q1$))" and "do(AS,inform($AH, Q2$))" can occur simultaneously only if $Q1$ and $Q2$ are related in this way. This latter implication acts as, essentially, a unique names axiom over inform acts; if it is not shared knowledge that $Q1$ is the same as $Q2$ then the act of communicating $Q1$ is different from the act of communicating $Q2$, since they may have different consequences

For example, if Jack and Jane share the knowledge that George Bush is the President and that 1600 Pennsylvania Avenue is the White House, then the action of Jack informing Jane that Bush is at the White House is identical to the act of Jack informing Jane that the President is at 1600 Pennsylvania Avenue. If they do not share this knowledge, then these two acts are different.

occurs(do(AS,inform($AH, Q1$)),$S1, S2$) $\Rightarrow$
[occurs(do(AS,inform($AH, Q2$)),$S1, S2$) $\Leftrightarrow$
[$\forall_{S1A}$ sk_acc($AS, AH, S1, S1A$) $\Rightarrow$
[holds($S1A, Q1$) $\Leftrightarrow$ holds($S1A, Q2$)]]]

I.5. The final axiom is a comprehension axiom schema, which states that any property of situations that can be stated in the language is a fluent.

Let $\mathcal{L}$ be a first-order language containing the primitives "<", "holds", "occurs", "do", "k_acc", "sk_acc", "communicate" and "inform" plus domain- and problem-specific primitives. Let $\alpha(S)$ be a formula in $\mathcal{L}$ with exactly one free variable S of sort "situation". (α may have other free variables of other sorts.) Then the closure of the following formula is an axiom:

$\exists_Q \forall_S$ holds(S, Q) $\Leftrightarrow \alpha(S)$.

(The closure of a formula β is β scoped by universal quantifications of all its free variables.)

Our theory does not include a frame axiom over knowledge. Informative actions cannot be the only knowledge-producing actions; if $A1$ does something that changes the preconditions for actions of $A2$, then $A2$ will become aware of the fact, if only because the space of feasible action changes. We have not found a correct formulation of the frame axiom that applies in general for this setting. (See (Davis, 1987) and (Scherl and Levesque, 2003) for theories that do use frame axioms over knowledge.) In any case, frame axioms over knowledge are often unimportant; in many applications, there is no need to establish that an agent will be ignorant of a given fact.

Sample Inferences

We illustrate the power of the above theory with two toy problems. First, we prove a useful lemma.

Lemma 1: If AS informs AH that Q, then, when the inform act is complete, AH knows that Q held when the inform act was begun.
occurs(do(AS,inform(AH, Q)),$S0, S1$) $\wedge$
k_acc($AH, S1, S1A$) $\Rightarrow$
$\exists_{S0A}$ occurs(do(AS,inform(AH, Q)), $S0A, S1A$) $\wedge$
holds($S0A, Q$).

Proof:
Let as, ah, q, s0, s1, s1a satisfy the left side of the above implication.
By K.7, sk_acc(as,ah,s1,s1a).
By I.3 there exists s0a such that
occurs(do(as,inform(ah,q)),s0a,s1a).
By K.1, k_acc(as,s0a,s0a).
By I.2, holds(s0a,q).

Sample Inference 1: Given:

X.1 Sam knows in s0 that it will be sunny on July 4.
[k_acc(sam,s0,$S0A$) $\wedge$ $S0A < S1A$ $\wedge$
time($S1A$)=july4] $\Rightarrow$
holds($S1A$,sunny).

X.2 In any situation, if it is sunny, then Bob can play tennis.
$\forall_S$ holds(S,sunny) $\Rightarrow \exists_{S1}$ occurs(do(bob,tennis),$S, S1$)

X.3 Sam can always communicate with Bob.
$\forall_{S1} \exists_{S2}$ occurs(do(sam,communicate(bob)),$S1, S2$).

Infer:

X.P Sam knows that there is an action he can do (e.g. tell Bob that it will be sunny) that will cause Bob to know that he will be able to play tennis on July 4.

k_acc(sam,s0,$S0A$) $\Rightarrow$
$\exists_{Z,S1A}$ occurs(do(sam,Z),$S0A, S1A$) $\wedge$
$\forall_{S2A,S2B}$ [occurs(do(sam,Z),$S0A, S2A$) $\wedge$
k_acc(bob,$S2A, S2B$) $\wedge$
$S2B < S3B$ $\wedge$ time($S3B$)=july4] $\Rightarrow$
$\exists_{S4B}$ occurs(do(bob,tennis),$S3B, S4B$).

Proof:
By the comprehension axiom I.5 there is a fluent q1 that holds in any situation S just if it will be sunny on July 4 following S.
holds(S,q1) $\Leftrightarrow$
[$\forall_{S1}$ [$S < S1$ $\wedge$ time($S1$)=july4] $\Rightarrow$ holds($S1$,sunny)].
Let z1=inform(bob,q1). By axioms I.2, X.1, and X.3, do(sam,z1) is feasible in s0;
$\exists_{S1}$ occurs(do(sam,z1),s0,$S1$).
By axiom K.5, Sam knows in s0 that do(sam,z1) is feasible.
$\forall_{S0A}$ k_acc(s0,$S0A$) $\Rightarrow$
$\exists_{S1A}$ occurs(do(sam,z1),$S0A, S1A$).
Let s0a be any situation such that k_acc(sam,s0,s0a).
Let s2a be any situation such that
occurs(do(sam,z1),s0a,s2a),
Let s2b be any situation such that k_acc(bob,s2a,s2b).
By Lemma 1, there exists s1b such that
occurs(do(sam,z1),s1b,s2b) and holds(s1b,q1).
Let s3b be any situation such that s2b < s3b and
time(s3b)=july4.
By T.8 and T.4, s1b < s3b.
By definition of q1, holds(s3b,sunny).
By X.2, there exists s4b such that
occurs(do(bob,tennis),s3b,s4b).
Applying the appropriate universal and existential abstractions over these constant symbols gives us formula X.P.

Sample Inference 2:
Given: Bob tells Alice that he has cheated on her. Alice responds by telling Bob that he has never told her anything she did not already know.

Infer: Bob now knows that Alice knew before he spoke that he had cheated on her.

Note that the inference only works if the two are speaking; if they are communicating by mail, then Bob may consider it possible that Alice sent her letter before receiving his, in which case she would not be including his latest communication. Therefore to represent this inference, we add two new actions: "do(AS,speak(AH,Q))" is a special case of "do(AS,inform(AH,Q))"; and "do(AH,listen(AS))" is an action that always (ideally) takes place simultaneously with "do(AS,speak(AH,Q)). The function "listen" does not take a content as argument, because the hearer does not know the content until the communication is finished.

Y.1 Bob confesses to Alice that he has cheated on her.
$\exists_Q$ occurs(do(bob,speak(alice,Q)),s0,s1) $\wedge$
$\forall_S$ holds$(S,Q) \Leftrightarrow$
$\exists_{S2,S3}$ $S3 < S \wedge$ occurs(do(bob,cheat),$S2,S3$).

Y.2 Alice responds that Bob has never told her anything she didn't already know. (Equivalently, whenever he has told her anything, she already knew it.)
$\exists_Q$ occurs(do(alice,speak(bob,Q)),s1,s2) $\wedge$
$\forall_S$ holds$(S,Q) \Leftrightarrow$
$\forall_{S3,S4,Q1}$
$[S3 < S4 \leq S \wedge$
occurs(do(bob,inform(alice,$Q1$)),$S3,S4$)] $\Rightarrow$
$\forall_{S3A}$ k_acc(alice,$S3,S3A$) $\Rightarrow$ holds$(S3A,Q1)$.

Y.3 If AS speaks Q to AH, then AS informs AH of Q.
occurs(do(AS,speak(AH,Q)),$S1,S2$) $\Rightarrow$
occurs(do(AS,inform(AH,Q)),$S1,S2$).

Y.4 If AS speaks Q to AH, then AH concurrently listens to AS.
[$\exists_Q$occurs(do(AS,speak(AH,Q)),$S1,S2$)] $\Leftrightarrow$
occurs(do(AH,listen(AS)),$S1,S2$)

Y.5 A speaker can only say one thing at a time.
[occurs(do(AS,speak($AH1,Q1$)),$S1,S2$) $\wedge$
occurs(do(AS,speak($AH2,Q2$)),$S3,S4$) $\wedge$
$S1 < S4 \wedge S3 < S2] \Rightarrow$
$[Q1 = Q2 \ \wedge\ S1 = S3\ \wedge S2 = S4]$

Infer:

Y.P Bob now knows that Alice had already known, before he spoke, that he had cheated on her.
$\forall_{S2A}$ k_acc(bob,s2,$S2A$) $\Rightarrow$
$\exists_{S0A,S1A,Q1}$ $S1A < S2A \wedge$
occurs(do(bob,inform(alice,$Q1$)),$S0A,S1A$) $\wedge$
[$\forall_{S0B}$ k_acc(alice,$S0A,S0B$) $\Rightarrow$
$\exists_{S3B,S4B}$ $S4B < S0B \wedge$
occurs(do(bob,cheat),$S3B,S4B$)].

Proof: Let q1 be the content of Bob's statement in Y.1, and let q2 be the content of Alice's statement in Y.2.
By K.4 and Y.3, Bob knows in s2 that he has informed Alice of q1.
$\forall_{S2A}$ k_acc(bob,s2,$S2A$) $\Rightarrow$
$\exists_{S0A,S1A}$ $S1A < S2A \wedge$
occurs(do(bob,inform(alice,q1)),$S0A,S1A$).

By Lemma 1, Bob knows in s2 that q2 held before Alice's speech act.
$\forall_{S2A}$ k_acc(bob,s2,$S2A$) $\Rightarrow$
$\exists_{S1A}$ occurs(do(alice,speak(bob,q2)),$S1A,S2A$) $\wedge$
holds($S1A$,q2).

Let s2a be any situation such that k_acc(bob,s2,s2a), and let s1a be a corresponding value of $S1A$ satisfying the above formula. By Y.4, Bob listened while Alice spoke.
occurs(do(bob,listen(alice)),s1a,s2a).
occurs(do(bob,listen(alice)),s1,s2).
By K.5, there exists an s11a such that k_acc(bob,s1,s11a) and occurs(do(bob,listen(alice)),s11a,s2a).
By Y.4, Alice must have spoken something from s11a to s2a.
By T.5 s11a and s1a are ordered.
By Y.5, s11a=s1a.
Thus, holds(s1a,q2); in other words, by definition of q2
(YY) $\forall_{S3,S4,Q1}$
$[S3 < S4 \leq$ s1 $\wedge$
occurs(do(bob,inform(alice,$Q1$)),$S3,S4$)] $\Rightarrow$
$\forall_{S3A}$ k_acc(alice,$S3,S3A$) $\Rightarrow$ holds$(S3A,Q1)$.
By K.4 and Y.3, Bob knows in s1 that he has informed Alice of q1.
$\forall_{S1A}$ k_acc(bob,s1,$S1A$) $\Rightarrow$
$\exists_{S0A}$ occurs(do(bob,inform(alice,q1)),$S0A,S1A$).
In particular, therefore,
$\exists_{S0A}$ occurs(do(bob,inform(alice,q1)),$S0A$,s1a).
Let s0a be a situation satisfying the above. Combining this with formula (YY) above gives
$\forall_{S0B}$ k_acc(alice,s0a,$S0B$) $\Rightarrow$ holds($S0B$,q1).
Applying the definition of q1, we get the desired result.

Paradox

The following Russell-like paradox seems to threaten our theory:

Paradox: Let Q be a fluent. Suppose that over interval $[S0,S1]$, agent a1 carries out the action of informing a2 that Q holds. Necessarily, Q must hold in $S0$, since agents are not allowed to lie (axiom I.2). Let us say that this communication is *immediately obsolete* if Q no longer holds in $S1$. For example, if it is raining in s0, the event of a1 telling a2 that it is raining occurs over [s0,s1], and it has stopped raining in s1, then this communication is immediately obsolete. Now let us say that situation S is "misled" if it is the end of an immediately obsolete communication. As being misled is a property of a situation, it should be definable as a fluent. Symbolically,

holds(S,misled) $\equiv$
$\exists_{Q,A1,A2}$ occurs(do($A1$,inform($A2,Q$)),$S0,S$) $\wedge$
$\neg$holds(Q,S)

Now, suppose that, as above, in s0 it is raining; from s0 to s1, a1 tells a2 that it is raining; and in s1 it is no longer raining and a1 knows that it is no longer raining. Then a1 knows that "misled" holds in s1. Therefore, (axiom I.2) it is feasible for a1 to tell a2 that "misled" holds in s1. Suppose that, from s1 to s2, the event occurs of a1 informing a2 that "misled" holds. The question is now, does "misled" hold in s2? Well, if it does, then what was communicated over [s1,s2] still holds in s2, so "misled" does not hold; but if it doesn't, then what was communicated no longer holds, so "misled" does hold in s2.

The flaw in this argument is that it presumes a unique names assumption that we have explicitly denied in axiom I.4. The argument presumes that if fluent $Q1 \neq Q2$, and do($A1$,inform($A2, Q1, T$)) occurs from s1 to s2, then do($A1$,inform($A2, Q2, T$)) does not occur. (Our English description of the argument used the phrase "what was communicated between s1 and s2", which presupposes that there was a unique content that was communicated.) But axiom I.4 asserts that many different fluents are communicated in the same act. Therefore, the argument collapses.

In particular, suppose that there is some fluent $\Delta(S)$ such that a1 and a2 have shared knowledge that Δ holds in s1 but not in s2. For instance, if a1 and a2 have shared knowledge that the time is 9:00 AM exactly, then $\Delta(S)$ could be "The time of S is 9:00 AM." Now, let q1 be any fluent, and suppose that occurs(do(a1,inform(a2,q1)),s1,s2). Let q2 be the fluent defined by the formula

$\forall_S$ holds(S,q2) $\Leftrightarrow$ holds(S,q1) $\wedge$ $\Delta(S)$.

By assumption, it is shared knowledge between a1 and a2 that holds(s1,q2) $\Leftrightarrow$ holds(s1,q1). Hence, by axiom I.4, occurs(do(a1,inform(a2,q2)),s1,s2). But by construction q2 does not hold in s1; hence the occurrence of do(a1,inform(as,q2)) from s1 to s2 is immediately obsolete. Therefore "misled" holds following *any* informative act.

Changing the definition of misled to use the universal quantifier, thus:

holds(S,misled) $\equiv$
$\forall_{Q,A1,A2}$ occurs(do($A1$,inform($A2, Q$)),$S0, S$) $\wedge$
$\neg$holds(Q, S)

does not rescue the contradiction. One need only change the definition of q2 above to be

$\forall_S$ holds(S,q2) $\Leftrightarrow$ holds(S,q1) $\vee$ $\neg\Delta(S)$.

Clearly, the new definition of "misled" *never* holds after any informative act.

Of course, if we extend the theory to include the underlying locutionary act, then this paradox may well return, as the locutionary act that occurs presumably is unique. However, as the content of a locutionary act is a quoted string, we can expect to have our hands full of paradoxes in that theory; this "misled" paradox will not be our biggest problem (Morgenstern,1988).

Unexpected Hanging

The well-known paradox of the unexpected hanging (also known as the surprise examination) (Gardner, 1991; Quine, 1953) can be formally expressed in our theory; however, the paradox does not render the theory inconsistent. (The analysis below is certainly *not* a philosophically adequate solution to the paradox, merely an explanation of how our particular theory manages to side-step it.)

The paradox can be stated as follows:

> A judge announces to a prisoner, "You will be hung at noon within 30 days; however, that morning you will not know that you will be hung that day." The prisoner reasons to himself, "If they leave me alive until the 30th day, then I will know that morning that they will hang me that day. Therefore, they will have to kill me no later than the 29th day. So if I find myself alive on the morning of the 29th day, I can be sure that I will be hung that day. So they will have to kill me no later than the 28th day ... So they can't kill me at all!"
>
> On the 17th day, they hung him at noon. He did not know that morning that he would be hung that day.

We can express the judge's statement as follows:

occurs(do(judge,inform(prisoner,Q)),s0,s1) $\wedge$
$\forall_S$ holds(S, Q) $\Leftrightarrow$
$\forall_{SX}$ [$S < SX$ $\wedge$ date(SX) = date(S)+31] $\Rightarrow$
$\exists_{SH,SM,SMA,SHA}$
$S < SM < SH < SX$ $\wedge$ hour(SH)=noon $\wedge$
holds(SH,hanging) $\wedge$ hour(SM)=9am $\wedge$
date(SM)=date(SH) $\wedge$
k_acc(prisoner,SM, SMA) $\wedge$ $SMA < SHA$ $\wedge$
hour(SHA)=noon $\wedge$ date(SM)=date(SH) $\wedge$
$\neg$holds(SHA,hanging).

That is: the content of the judge's statement is the fluent defined by the following formula over S: On any timeline starting in S and going through some SX 31 days later, there is a situation SH at noon where you will be hung, but that morning SM you will not know you will be hung; that is, there is a SMA knowledge accessible from SM which is followed at noon by a situation SHA in which you are not hung.

Let UH^{lang} be the judge's statement in English and let UH^{logic} be the fluent defined in the above formula. Let "kill(K)" be the proposition that the prisoner will be killed no later than the Kth day, and let "kill_today" be the fluent that the prisoner will be killed today. It would appear that UH^{lang} is true; that the judge knows that in s0 that it is true, and that UH^{logic} means the same as UH^{lang}. By axiom I.2, if the judge knows that UH^{logic} holds in s0, then he can inform the prisoner of it. How, then, does our theory avoid contradiction?

The first thing to note is that the prisoner *cannot* know UH^{logic}. There is simply no possible worlds structure in which the prisoner knows UH^{logic}. The proof is exactly isomorphic to the sequence of reasoning that prisoner goes through. Therefore, by Lemma 1 above, the judge cannot inform the prisoner of UH^{logic}; if he did, the prisoner would know it to be true.

The critical point is that there is a subtle difference between UH^{lang} and UH^{logic}. The statement UH^{lang} asserts that the prisoner *will* not know kill_today — this means even *after* the judge finishes speaking. In our theory, however, one can only communicate properties of the situation at the beginning of the speech act and there is no way to refer to what *will* happens as distinguished from one *could* happen. So what UH^{logic} asserts is that the prisoner will not know kill_today *whatever* the judge decides to say or do in s0.

In fact, it is easily shown that either [the judge does not know in s0 that UH^{logic} is true], or [UH^{logic} is false]. It depends on what the judge knows in s0. Let us suppose that in s0, it is inevitable that the prisoner will be killed on day 17 (the executioner has gotten irrevocable orders.) There are two main cases to consider.

- Case 1: All the judge knows kill(K), for some $K > 17$. Then the most that the judge call tell the prisoner is kill(K). In this case, UH^{logic} is in fact true in s0, but the judge does not know that it is true, because as far as the judge knows, it is possible that (a) he will tell the prisoner kill(K) and (b) the prisoner will be left alive until the Kth day, in which case the prisoner would know kill_today on the morning of the Kth day.

- Case 2: The judge knows kill(17). In that case, UH^{logic} is not even true in s0, because the judge has the option of telling the prisoner kill(17), in which case the prisoner will know kill_today on the morning of the 17th day.

Again, we do not claim that this is an adequate solution to the philosophical problem, merely an explanation of how our formal theory manages to remain consistent and side-step the paradox. In fact, in the broader context the solution is not at all satisfying, for reasons that may well become serious when the theory is extended to be more powerful. There are two objections. First, the solution depends critically on the restriction that agents cannot talk about what *will* happen as opposed to what *can* happen; in talking about the future, they cannot take into account their own decisions or commitments about what they themselves are planning to do. One can extend the outer theory so as to be able to *represent* what will happen — in (Davis and Morgenstern, 2004), we essentially do this — but then the comprehension axiom I.5 must be restricted so as to exclude this from the scope of fluents that can be the content of an "inform" act. We do not see how this limitation can be overcome.

The second objection is that it depends on the possibility of the judge telling the prisoner kill(17) if he knows this. Suppose that we eliminate this possibility? Consider the following scenario: The judge knows kill(17), but he is unable to speak directly to the prisoner. Rather, he has the option of playing one of two tape recordings; one says "kill(30)" and the other says UH^{logic}. Now the theory is indeed inconsistent. Since the prisoner cannot know UH^{logic} it follows that the judge cannot inform him of UH^{logic}; therefore the only thing that the judge can say is "kill(30)". But in that case, the formula "UH^{logic}" is indeed true, and the judge knows it, so he should be able to push that button.

To axiomatize this situation we must change axiom I.2 to assert that that the only possible inform acts are kill(30) and UH^{logic}.

Within the context of our theory, it seems to me that the correct answer is "So what?" Yes, you can set up a Rube Goldberg mechanism that creates this contradiction, but the problem is not with the theory, it is with the axiom that states that only these two inform acts are physically possible.

In a wider context, though, this answer will not serve. After all, it is physically possible to create this situation, and in a sufficiently rich theory of communication, it will be provable that you can create this situation. However, such a theory describing the physical reality of communication must include a theory of locutionary acts; i.e. sending signals of quoted strings. As mentioned above such a theory will run into *many* paradoxes; this one is probably not the most troublesome.

Consistency

Two paradoxes have come up, but the theory has side-stepped them both. How do we know that the next paradox won't uncover an actual inconsistency in the theory? We can eliminate all worry about paradoxes once and for all by proving that the theory is consistent. We do this by constructing a model satisfying the theory. More precisely, we construct a fairly broad class of models, establishing (informally) that the theory is not only consistent but does not necessitate any weird or highly restrictive consequences. (Just showing soundness with respect to a model or even completeness is not sufficient for this. For instance, if the theory were consistent only with a model in which every agent was always omniscient, and inform acts were therefore no-ops, then the theory would be *consistent* but not of any interest.)

As usual, establishing soundness has three steps: defining a model, defining an interpretation of the symbols in the model, and establishing that the axioms are true under the interpretation.

Our class of models is (apparently) more restrictive than the theory;[1] that is, the theory is not complete with respect to this class of models. The major additional restrictions in our model are:

I. Time must be discrete. We believe that this restriction can be lifted with minor modifications to the axioms, but this is beyond the scope of this paper. We hope to address it in future work.

II. Time must have a starting point; it cannot extend infinitely far back. It would seem to be very difficult to modify our proof to remove this constraint; at the current time, it seems to depend on the existence of highly non-standard models of set theory.

III. A knowledge accessibility link always connects two situations whose time is equal, where "time" measure the number of clock ticks since the start. In other words, all agents always have common knowledge of the time. In a discrete structure, this is a consequence of the axiom of memory. Therefore, it is not, strictly speaking, an additional restriction; rather, it is a non-obvious consequence of restriction (I). If we extend the construction to a non-discrete time line, some version of this restriction must be stated separately.

There are also more minor restrictions; for example, we will define shared knowledge to be the true transitive closure of knowledge, which is not expressible in a first-order language.

Theorem 1 below states that the axioms in this theory are consistent with essentially any physical theory that has a model over discrete time with a starting point state and physical actions to knowledge.

Definition 1: A *physical language* is a first-order language containing the sorts "situations,", "agents", "physical actionals", "physical actions", "physical fluents", and "clock

[1]The only way to be sure that the theory is more general than the class of models is to prove that it is consistent with a broader class of models.

times"; containing the non-logical symbols, "$<$", "do", "occurs", "holds", "time", and "communicate"; and excluding the symbols, "k_acc", "inform", and "sk_acc".

Definition 2: Let $\mathcal{L}$ be a physical language, let $\mathcal{T}$ be a theory over $\mathcal{L}$. $\mathcal{T}$ is an *acceptable physical theory* (i.e. acceptable for use in theorem 1 below) if there exists a model $\mathcal{M}$ and an interpretation $\mathcal{I}$ of $\mathcal{L}$ over $\mathcal{M}$ such that the following conditions are satisfied:

1. $\mathcal{I}$ maps the sort of clock times to the positive integers, and the relation $T1 < T2$ on clock times to the usual ordering on integers.
2. Axioms T.1 — T.9 in table 1 are true in $\mathcal{M}$ under $\mathcal{I}$.
3. Theory $\mathcal{T}$ is true in $\mathcal{M}$ under $\mathcal{I}$.
4. The theory is consistent with the following constraint: In any situation S, if any communication act is feasible, then arbitrarily many physically indistinguishable communication acts are feasible. This constraint can be stated in a first order axiom schema, which we here omit.

Condition (4) seem strange and hard to interpret, but in fact most reasonable physical theories satisfy it, or can be modified without substantive change to satisfy it.

Theorem 1: Let $\mathcal{T}$ be an acceptable physical theory, and let $\mathcal{U}$ be $\mathcal{T}$ together with axioms K.1 — K.7 and I.1 — I.5. Then $\mathcal{U}$ is consistent.

Sketch of proof:
The main sticking point of the proof is as follows: In order to satisfy the comprehension axiom, we must define a fluent to be any set of situations. However, if Q is a fluent, then the act of AS informing informing AH of Q in $S1$ generates a new situation; and if we generate a separate "inform" act for each fluent, then we would have a unsolvable vicious circularity.

Restriction (III) and axiom I.4 rescue us here. Let q1 be any fluent that holds in situation s1. By axiom I.4, we can identify the act of AS informing AH of q1 starting in s1 with the act of AS informing AH of any other q2, such that AS and AH jointly know in s1 that q1 iff q2. Let t1=time(s1). By condition (III), in s1, AS and AH jointly know that the current time is t1. Let q2 be the fluent such that holds(S,q2) $\Leftrightarrow$ holds(S,q1) $\wedge$ time(S)=t1. Then AS and AH have shared knowledge in s1 that q1 is equivalent to q2. Applying this reasoning generally, it follows that the content of an inform act need not be a *general* set of situations, only a set of situations contemporaneous with the start of the inform act. This limitation allow us to break the circularity in the construction of situations and informative acts: the content of informative acts starting at time K is a subset of the situations whose time is K; informative acts starting in time K generate situations whose time is $K+1$.

Therefore, we can use the "algorithm" shown in table 3 to construct a model of the theory $\mathcal{U}$.

Once the model has been constructed, defining the interpretation and checking that the axioms are valid is straightforward. The only part that requires work is establishing

Constructing a model

```
Let M be a collection of branching time models of
   theory T;
Create a set of initial situations at time 0.
   Map each initial situation S to an initial situation
         PHYS(S) in M.
for (each agent A), define the relation K_ACC(A, ·, ·)
   to be some equivalent relation over the initial situations.
for (K=0 to ∞) do {
   for (each situation S of time K) do {
         for (each physical state PS following PHYS(S) in M)
           construct a new situation S1 and mark PHYS(S1)=PS;
         for (each pair of agents AS,AH) do {
           if (in M there is an act starting in S of AS
               communicating to AH)
           then {
             SSL := the set of situations knowledge
                         accessible from S relative to AS;
             SSU := the set of situations accessible
                         from S relative to the shared
                         knowledge of AS and AH;
             for (each set SS that is a subset of SSU
                 and a superset of SSL) do {
               construct an action "inform(AS,AH,SS)"
                 starting in S;
               construct a successor S1 of S corresponding
                 to the execution of this action;
               label PHYS(S1) to be a physical state in M
                 following a communicate action in PHYS(S);
             }
           } } }
         use the axioms of knowledge to construct a valid set of
           knowledge accessibility relations over the new situations
}
```

Table 3: Construction of a model

that the new model still satisfies the physical theory. This follows from condition (4) on the theory $\mathcal{T}$; since there can exist arbitrarily many communicative acts in any situation, the addition of a bunch more, in the form of new "inform" acts, cannot be detected by the first-order theory $\mathcal{T}$.

The full details of the proof can be found in an appendix to this paper, at the URL http://cs.nyu.edu/faculty/davise/kr04-appendix.ps and .pdf.

It is possible to strengthen theorem 1 to add to $\mathcal{U}$:

- Any specification of knowledge and ignorance at time 0, subject to a few conditions relating these to $\mathcal{T}$ (e.g. that these specifications and $\mathcal{T}$ cannot require a incompatible numbers of agents);
- Axioms specifying that specified physical actions or situations cause an agent to gain knowledge.

However, the correct statement of the theorem becomes complex. Again, see the appendix.

Related Work

The theory presented here was originally developed as part of a larger theory of multi-agent planning (Davis and Morgenstern, 2004). That theory includes requests as speech acts as well as informative speech acts. However, our analysis of informative acts there was not as deep or as extensive in scope.

As far as we know, this is the first attempt to characterize the content of communication as a first-order property of possible worlds. Morgenstern (1988) develops a theory in which the content of communication is a string of characters. A number of BDI models incorporate various types of communication. The general BDI model was first proposed by Cohen and Perrault (1979); within that model, they formalized illocutionary acts such as "Request" and "Inform" and perlocutionary acts such as "Convince" using a STRIPS-like representation of preconditions and effects on the mental states of the speaker and hearer. Cohen and Levesque (1990) extend and generalize this work using an full modal logic of time and propositional attitudes. Here, speech acts are *defined* in terms of their effects; a request, for example, is any sequence of actions that achieves the specified effect in the mental state of the hearer.

Update logic (e.g. Plaza 1989; van Benthem 2003) combines dynamic logic with epistemic logic, introducing the dynamic operator $[A!]\phi$, meaning "ϕ holds after A has been truthfully announced.". The properties of this logic have been extensively studied. Baltag, Moss, and Solecki (2002) extend this logic to allow communication to a subset of agents, and to allow "suspicious" agents. Colombetti (1999) proposes a *timeless* modal language of communication, to deal with the interaction of intention and knowledge in communication. Parikh and Ramanujam (2003) present a theory of *messages* in which the meaning of a message is interpreted relative to a protocol.

There is a large literature on the applications of modal logics of knowledge to a multi-agent systems. For example, Sadek et al. (1997) present a first-order theory with two modal operators $\mathrm{B}_i(\phi)$ and $\mathrm{I}_i(\phi)$ meaning "Agent i believes that ϕ" and "Agent i intends that ϕ" respectively. An inference engine has been developed for this theory, and there is an application to automated telephone dialogue that uses the inference engine to choose appropriate responses to requests for information. However, the temporal language associated with this theory is both limited and awkward; it seems unlikely that the theory could be applied to problems involving multi-step planning. (The dialogue application requires only an immediate response to a query.)

The multi-agent communication languages KQML (Finin et al., 1993) and FIPA (FIPA, 2001) provide rich sets of communication "performatives". KQML was never tightly defined (Woolridge 2002.) FIPA has a formal semantics defined in terms of the theory of (Sadek et al. 1997) discussed above. However, the content of messages is unconstrained; thus, the semantics of the representation is not inherently connected with the semantics of the content, as in our theory.

Other modal theories of communication, mostly propositional rather than first-order, are discussed in (Wooldridge and Lomuscio, 2000; Lomuscio and Ryan, 2000; Rao, 1995).

Conclusions

We have developed a theory of communications which allows the content of an informative act to include quantifiers and logical operators and to refer to physical states, events including other informative acts, and states of knowledge; all these in the past, present, or possible futures. We have proven that this theory is consistent, and compatible with a wide range of physical theories. We have examined how the theory avoids two potential paradoxes, and discussed how these paradoxes may pose a danger when these theories are extended. Elsewhere (Davis and Morgenstern, 2004) we have shown that the theory can be integrated with a similarly expressive theory of multi-agent planning.

The most important problems to be addressed next are:

- Replacing the explicit manipulation of possible worlds and knowledge accessibility relations with some more natural representation, such as modal operators.
- Continuing our work on integrating this theory of communication with a theory of planning.
- Extending the theory to allow continuous time line.
- Integrating a theory of locutionary acts (Morgenstern, 1988).

References

Baltag, A., Moss, L. and Solecki, S. 2002. "The Logic of Public Announcements: Common Knowledge and Private Suspicions."

Benthem, J. van. 2003. " 'One is a Lonely Number': on the logic of communication." ILLC Tech Report 2003-07, Institute for Logic, Language and Computation, University of Amsterdam.

Cohen, P.R. and Perrault, C.R. 1979."Elements of a plan-based theory of speech acts." *Cognitive Science,* vol. 3, no. 3, pp. 177-212.

Cohen, P.R. and Levesque, H. 1990. “Intention is choice with commitment” *Artificial Intelligence*, vol. 42, nos. 2-3, pp. 213-261.

Colombetti, M. 1999. “A Modal Logic of Intentional Communication.”*Mathematical Social Sciences*, vol. 38, pp. 171-196.

Davis, E. 1988. “Inferring Ignorance from the Locality of Visual Perception.” *Proc. AAAI-88*, pp. 786-790

Davis, E. and Morgenstern, L. 2004. “ A First-Order Theory of Communication and Multi-Agent Plans.” Submitted to *Journal of Logic and Computation.*

Finin et al. 1993. “Specification of the KQML agent communication language.” DARPA knowledge sharing initiative external interfaces working group.

FIPA 2001. “The foundation for intelligent physical agents.” http://www.fipa.org/

Gardner, M. 1991. *The Unexpected Hanging and Other Mathematical Diversions.* Chicago University Press.

Lomuscio, A. and Ryan, M. 2000. “A spectrum of modes of knowledge sharing between agents.” *Intelligent Agents VI: Agent Theories, Architectures, and Languages.* Lecture Notes in Artificial Intelligence 1757, Springer-Verlag, pp. 13-26.

McDermott, D. 1982. “A Temporal Logic for Reasoning about Processes and Plans.” *Cognitive Science*, Vol. 6, pp. 101-155.

Moore, R. 1980. “Reasoning about Knowledge and Action.” Tech. Note 191, SRI International, Menlo Park, CA.

Moore, R. 1985. “A Formal Theory of Knowledge and Action.” In Jerry Hobbs and Robert Moore, (eds.) *Formal Theories of the Commonsense World.* ABLEX Publishing, Norwood, New Jersey, pp. 319-358.

Morgenstern, L. 1987. “Foundations of a Logic of Knowledge, Action, and Communication.” NYU Ph.D. Thesis.

Parikh, R. and Ramanujam, R., 2003. “A Knowledge-Based Semantics of Messages.” *Journal of Logic, Language, and Information.* vol. 12 no. 4.

Plaza, J. 1989. “Logics of Public Announcements.” *Proc. 4th International Symposium on Methodologies for Intelligence Systems.*

Quine, W.V.O. 1953. “On a So-Called Paradox.” *Mind* vol. 62, pp. 65-67.

Rao, A.S. 1995. “Decision Procedures for Propositional Linear Time Belief-Desire-Intention Logics.” *Intelligent Agents II: Agent Theories, Architectures, and Languages.* Lecture Notes in Artificial Intelligence 1037, Springer-Verlag, pp. 33-48.

Reiter, R. 2001. *Knowledge in Action: Logical Foundations for Specifying and Implementing Dynamical Systems.* MIT Press.

Sadek, M.D., Bretier, P. and Panaget, F. 1997. “ARTIMIS: Natural dialogue meets rational agency,] *Proc. IJCAI-97*, pp. 1030-1035.

Scherl, R. and Levesque, H. 1993. “The Frame Problem and Knowledge Producing Actions.” *Proc. AAAI-93*, pp. 689-695.

Scherl, R. and Levesque, H. 2003. “Knowledge, action, and the frame problem.” *Artificial Intelligence*, vol. 144 no. 1, pp. 1-39.

Woolridge, M. 2002. *An Introduction to MultiAgent Systems.* John Wiley and Sons.

Wooldridge, M. and Lomuscio, A. 2000. “Reasoning about Visibility, Perception, and Knowledge.” *Intelligent Agents VI: Agent Theories, Architectures, and Languages.* Lecture Notes in Artificial Intelligence 1757, Springer-Verlag, pp. 1-12.