Knowledge and implicit knowledge in a distributed environment: Preliminary Report

Ronald Fagin
Moshe Y. Vardi

IBM Almaden Research Center
650 Harry Road
San Jose, California 95120-6099

Abstract: We characterize the states of knowledge that are attainable in distributed systems, where communication is done by unreliable message exchange. The reason that certain states of knowledge are unattainable is a conservation principle which says that information about "nature" that can be obtained by combining all of the knowledge of the members of a closed system is preserved. We axiomatize the class of formulas in the propositional modal logic of knowledge that are valid in attainable knowledge states, and we determine the complexity of the decision problem.

Appeared in Conference on Theoretical Aspects of Reasoning about Knowledge, ed. J. Y. Halpern, Morgan Kaufmann, 1986, pp. 187-206.

1. Introduction

A recent and exciting paradigm in the area of distributed systems, first put forward by Halpern and Moses [HM1], is that the right way to understand distributed protocols is by considering how communication changes the state of *knowledge* of distribuied processes. To quote from [HM1], "Many tasks in a distributed system directly involve the achievement of specific states of knowledge, and others crucially depend on a variety of constraints on the state of knowledge of the parties involved". This paradigm has inspired computer scientists (cf. [CM, FHV1, Leh, Pa, PR]) to study an area that has so far been in the realm of economics [Au], philosophy [Hi], and artificial intelligence [MH] - the *logic of knowledge.*

In order to formalize reasoning about knowledge, we need semantic models for knowledge. The most common approach to modelling knowledge, due to Hintikka [Hi], is based on the *possible worlds semantics.* In this approach, the information that a "player" (or "agent" or "process") has about the world may be incomplete; rather than knowing precisely what the actual state of the world is, the player may only know that the actual state of the world belongs to a given set of possible states (the so-called possible worlds). A player then *knows* a fact φ to be true if φ is true in all the states that the player thinks are possible. Possible world semantics has been formalized using either Kripke structures [Kr] or modal structures [FV]. When used to model knowledge, modal structures are called *knowledge structures* [FHV1].

We can use these semantic models for knowledge to interpret formulas in the logic of knowledge. These formulas are propositional modal formulas, where for every player i we have a modality K_i. Intuitively, the formula $K_i\varphi$ says "player i knows φ". In order to understand the nature of knowledge better, it is helpful to characterize knowledge by axiomatizing valid formulas (the formulas that are satisfied by all knowledge structures). It turns out that the well-known modal logic S5 (which is described in the body of the paper) is a sound and complete axiomatization for knowledge structures, which may suggest that S5 is an appropriate formalism for reasoning about knowledge in distributed systems.

Knowledge structures can be viewed as abstract models for knowledge. Namely, they model all possible states of knowledge with no concern as to how knowledge is acquired in the first place. To reason formally about knowledge in distributed systems, we need, however, to know which states of knowledge are attainable in such systems. In particular, since players in distributed systems communicate with each other exclusively by exchanging messages, we need to know what states of knowledge are attainable via such communication.

To this end we start with a concrete model of knowledge. The basic element in this model is a *run.* A run is a description of a distributed system over time; it consists of a description of the "real world" or "nature", which we assume does not change as a result of communication in the system, the players' initial information about nature, and the messages sent and received by the players. Two runs are *equivalent* with respect to a player i, or "i-equivalent", if they are indiscernible as far as player i is concerned. A player i is said to "know" φ in a run S if φ is true in all runs that are i-equivalent to S. (This concrete model of knowledge is suggested in [CM, DFIL, HF, HM1, PR, RP], and is also implicit in [Dw].) It is not hard to verify that under this interpretation of "know", the axiom system S5 that we have discussed is sound. That is, the axioms and rules of inference of S5 all hold under this interpretation.

We now consider the abstract counterpart of the above concrete model, i.e., the knowledge structures that correspond to the run-based model. It turns out that we do *not* get *all* knowledge structures. In other words, there are knowledge structures that describe knowledge states that are unattainable via message exchange. In particular, S5 is not complete for reasoning about knowledge in distributed systems! For example, if the only primitive proposition is p and if the only players are 1 and 2, then the formula

(1) $K_1((p \wedge \sim K_1 p \wedge \sim K_2 p) \vee (\sim p \wedge \sim K_1 \sim p \wedge K_2 \sim p))$

is not satisfiable in distributed systems, even though it is S5-consistent. To get a complete axiomatization it is necessary to augment S5 with an additional axiom. (We shall see that if there is another primitive proposition besides p or if there is another player besides players 1 and 2, then (1) *is* satisfiable in distributed systems.)

To better understand this phenomenon, it is useful to consider the logic of knowledge and *implicit* knowledge. Implicit knowledge, introduced by Halpern and Moses [HM1], is the knowledge that can be obtained by pooling together the knowledge of a group. Put differently, the implicit knowledge of a group G is what someone could infer given complete knowledge of what each member of G knows. For example, if Alice knows φ_1 and Bob knows $\varphi_1 \Rightarrow \varphi_2$, then together they have implicit knowledge of φ_2, even though neither of them might individually know φ_2. The basic feature of message-based knowledge is *conservation* of implicit knowledge of nature: that is, communication among the players cannot increase the implicit knowledge of the group as a whole about nature. This conservation principle is *dynamic*, in that it deals with *changes* in knowledge. Surprisingly, this dynamic principle has consequences on *static* implicit knowledge. Even more surprisingly, this principle not only affects what (static) implicit knowledge the players can have as a group but also what (static) knowledge individual players can have.

The main point that we are trying to get across in this paper is that knowledge in a distributed systems depends in a crucial way on the way in which processes communicate with each other. Here we investigate one particular model of communication, but this model is not more basic than other prevalent models. For example, in our model communication is unreliable. As we shall point out in the paper, if we assume that communication is reliable, than the effect on the attainable knowledge states is drastic. We believe that the issue of how communication affects knowledge deserves a great deal of further study.

The outline of this paper is as follows. In Section 2, we give the syntax and semantics of runs. In Section 3, we state and prove the conservation principle for implicit knowledge. In Section 4 we describe a property of implicit knowledge that follows from the conservation principle, and we give an axiom that captures this property. In Section 5, we discuss a concrete example which shows that S5 is not a complete axiomatization for communication-based knowledge. Specifically, we show that the formula (1) above is not satisfiable under communication. In Section 6, we give two sound and complete axiomatizations for knowledge under message exchange; one axiomatization involves implicit knowledge, and the other does not. In Section 7 we show that if we assume that communication is reliable, then the set of attainable knowledge states is restricted drastically. In Section 8, we discuss the effect of changing the class of messages. Sections 9-11 study the model theory of our framework. Section 9 reviews the definitions of knowledge worlds from [FHV1], and Section 10 characterizes those knowledge worlds that can arise under message exchange, which are called *message-based knowledge worlds*. In Section 11, we show that implicit knowledge behaves badly in general but nicely in message-based knowledge worlds. We conclude with some remarks in Section 12.

2. Runs

We assume that there is a fixed finite set of primitive propositions, and a fixed finite set $\mathcal{P}$ of players. The class of *formulas* is the smallest set that contains the primitive propositions, is closed under the Boolean connectives $\sim$ and $\wedge$, and contains $K_i\varphi$ if it contains φ, for each player i. The class of *extended formulas* is defined similarly, except that also $I\varphi$ is an extended formula if φ is. Thus, extended formulas allow also the modal operator I ("It is implicit knowledge that").

We are about to give the syntax and semantics of runs. Throughout this paper, we assume that communication is *synchronous* and proceeds in "rounds". We assume that messages may be lost, that is, never received. (As we shall show later, if messages are guaranteed to be delivered, then the situation changes radically.) We also assume that if a message is ever received, then it is received in the round it was sent.

We assume that some fixed truth assignment to the primitive propositions is "the actual truth assignment", or "nature". An alternative viewpoint (which is useful, for example, in statistics [Sa]) is that instead of primitive propositions and truth assignments, there is a fixed finite set of *primitive states*, and that "nature" is one of these primitive states. To make it easier to pass back and forth between these two viewpoints, we shall usually refer to a truth assignment as a primitive state. Let $\mathcal{N}$ be the set of primitive states.

We begin with an intuitive description of the "initial information" of each player, and how communication takes place. At the beginning (or "in the 0th round"), each player i is "told" a set $T(i)$ of primitive states, one of which is nature. We view $T(i)$ as player i's initial information about nature. In particular, if $T(i) = \{t\}$, where t is nature, then player i knows completely about nature, and if $T(i)$ is the set of all primitive states, then player i knows nothing about nature. One intuitive way to think about what we have just said is that before there is any communication between the players, each player "studies nature", and player i "learns" $T(i)$ (that is, player i gains the information that nature is a member of the set $T(i)$). No player has any information about any other player's initial information about nature. After players obtain this initial information about nature, all information is gained by messages that are sent between the players. Intuitively, no one ever "studies nature" again (we also assume that nature never changes). We make this assumption, since we are interested in characterizing the knowledge of each player when new information is gained *only* by message exchange. We leave as a problem how to characterize the knowledge of each player when it is possible for a player to gain directly more information about nature at any time.

In each round, each player may send any number of messages to the other players. For example, in round 3, player 1 may send three messages to player 2, no messages to player 3, and one message to player 4.

We now discuss the class $\mathcal{M}$ of messages. As we shall see later, in order to get our completeness results, the class $\mathcal{M}$ must be sufficiently rich; for example, the class of formulas (or even the class of extended formulas) is not sufficiently rich to serve as the class of messages. It is technically convenient to distinguish two types of messages: *messages about the past*, and *messages about the future*. "Messages about the past" talk about previous rounds; for example, on round 7, one message about the past is "I sent message φ to player j in round 5". "Messages about the future" make certain promises about future messages; for example, on round 7, one message about the future is "If ψ holds in round 31, then in round 31 the only messages I will send to player j are in the set Φ". We shall assume that every message is *honest* [HF]. In the case of messages about the past, honesty means that the message is true (our semantics is such that this will automatically guarantee that the sender of the message *knew* that the message was true when he sent it). In the case of messages about the future, honesty means that every promise is fulfilled. It is convenient to consider messages about the future to be honest at the time they are sent; however, the promises made must be kept. Thus, at a later date, messages about the future may be rendered dishonest. Shortly, we shall discuss other reasons why we distinguish between messages about the past and messages about the future. The class $\mathscr{M}$ of messages consists of the following (where φ is a message and Φ is a finite set of messages):

1. Messages about the past:
   a. "I knew θ just after round r", where θ is an extended formula.
   b. "I sent message φ to player j in round r".
   c. "I sent precisely the set Φ of messages to player j in round r".
   d. "I received message φ from player j in round r".
   e. "I received precisely the set Φ of messages from player j in round r".
   f. "Every message that I sent in round r to player j I still know to be true".
   g. Each finite Boolean combination of messages about the past.
2. Messages about the future:
   "If ψ holds in round r, then in round r the only messages I will send to player j are in the set Φ", where ψ is a message about the past.

Note that, for example, the message "I sent message φ to player j in round r" is a message about the past, even if φ is a message about the future. Note also that messages about the future simply restrict the class of future messages. In particular, sending no more messages at all automatically makes a message about the future honest.

In our examples, we often find it convenient to allow extended formulas as messages. If θ is an extended formula, and if the message θ is sent in round $r+1$, then this message θ can be viewed as a shorthand for the message "I knew θ just after round r".

Why are we so restrictive as to which messages about the future that we allow? First, it is shown in [HF] that serious problems arise if too general messages about the future are allowed. In particular, with more general messages about the future it is hard to make sense out of "honesty", and there does not seem to be a reasonable and natural semantics. Second, the messages we have defined are all we need in order for our results (in particular, our complete axiomatizations) to go through. Third, we shall show later (in Section 8) that in a certain sense, our results still hold if more messages are allowed; however, adding more messages can considerably complicate the semantics. Finally, with the class of messages we have defined, runs have the following nice property: if S is a k-round run, and if in the $(k+1)$st round no messages are sent (and, of course, none are received), then the result is a $(k+1)$-round run. In particular, every run is the prefix of an arbitrarily long run. If we were to allow messages about the future to be closed under Boolean combinations, then we would lose this property. For example, if a player were to send both a message about the future and its negation, then clearly there is no way to fulfill both of these "promises".

Why did we not define a "message about the past" by player i in round r to be simply an arbitrary disjunction of histories of player i up to round r (where a "history of player i up to round r" is a complete description of the set T of primitive states that he learned from nature in round 0, along with a complete description of the messages sent and received by player i in each round s where $1 \leq s < r$)? The reason is that the above messages would have to be infinite disjunctions, since there are an infinite number of messages (specifically, messages about the future) that player i could potentially send in, say, round 1.

We need a few more definitions before we can give the formal definition of the syntax and semantics of runs. If a message φ about the past is a Boolean combination of messages $\varphi_1, \ldots, \varphi_t$, each of one of the types (a) - (f), then we say that each φ_m $(1 \leq m \leq t)$ is a *direct submessage* of φ. If "I sent message ψ to player j in round r" is a direct submessage of φ, then let us say that φ *directly involves* r, and similarly for the other messages of types (a) - (f). For example, the message

"I knew θ just after round 2" $\vee$ "I sent message ψ to player j in round 5",

where ψ is the message "I sent message δ to player j in round 3", directly involves rounds 2 and 5 but does not directly involve round 3. Intuitively, φ directly involves r if round r is mentioned "at the top level" of φ. It is also convenient to say that the message about the future "If ψ holds in round r, then in round r the only messages I will send to player j are in the set Φ", *directly involves round r.*

We now begin the formal definition of the syntax and semantics of runs. A *k-round run* is a tuple $(\gamma, T, \text{sent}, \text{received})$, where (a) $\gamma \in \mathcal{W}$ (thus, γ is a primitive state); (b) T is a function T: $\mathcal{P} \rightarrow 2^{\mathcal{W}}$; and (c) sent is a function sent: $\mathcal{P} \times \mathcal{P} \times \{1, \ldots, k\} \rightarrow 2^{\mathcal{M}}$ (and similarly for received). Intuitively, γ is "nature"; $T(i)$ gives player i's "initial information about nature", as discussed earlier; $\text{sent}(i,j,r)$ is the set of messages sent by player i to player j in round r, and $\text{received}(i,j,r)$ is the set of messages received by player i from player j in round r. We assume that $\gamma \in T(i)$ for each player i (that is, nature is one of the possibilities that player i learns is possible in the 0th round). We also assume that $\text{sent}(i,j,0) = \varnothing = \text{received}(i,j,0)$ for each i,j (that is, no messages are sent or received in the 0th round). Intuitively, in the 0th round, players learn their initial information about nature, and in rounds 1,2,..., players communicate with each other. We also make the following assumptions:

1. $\text{received}(i,j,k+1) \subseteq \text{sent}(j,i,k+1)$ ("Each message is received in the round in which it was sent").
2. If φ is a message about the past that directly involves round r and $\varphi \in \text{sent}(i,j,k)$, then $r < k$ ("'Messages about the past' are really about the past").
3. If φ is a message about the future that directly involves round r, and if $\varphi \in \text{sent}(i,j,k)$, then $r > k$ ("'Messages about the future' are really about the future").

Finally, we wish to say that every message is honest. It is convenient for us to refer to an honest message sent by player i as *i-honest*. When we say that a message is *honest*, we mean that if it was sent by player i, then it is i-honest. To formally define what an i-honest message is, we assume inductively that we have completely defined k-round runs (in particular, we have defined honesty for k-round runs, and insisted that every message be honest in a k-round run). We then define what it means for a k-round run S to satisfy an extended formula φ (written $S \models \varphi$). We then define what it means for a message φ to be i-honest in a $(k+1)$-round run S (written $S \models_i \varphi$), and we then insist that every message in a $(k+1)$-round run be honest. The base case $(k=0)$ has been taken care of, since no messages are sent in the 0th round.

Let us say that the k-round runs $S = (\gamma, T, \text{sent}, \text{received})$ and $S' = (\gamma', T', \text{sent}', \text{received}')$ are *i-equivalent* (written $S \sim_i S'$) if the following conditions hold:

1. $T(i) = T'(i)$ ("player i receives the same information in the 0th round of both runs").
2. $\text{sent}(i,j,r) = \text{sent}'(i,j,r)$ for each player j and each round r with $1 \leq r \leq k$ ("player i sends the same messages to each player in the same rounds of both runs").
3. $\text{received}(i,j,r) = \text{received}'(i,j,r)$ for each player j and each round r with $1 \leq r \leq k$ ("player i receives the same messages from each player in the same rounds of both runs").

Thus, player i cannot distinguish between two i-equivalent k-round runs. We may say that in run S, player i thinks run S' is *possible* if $S \sim_i S'$.

We now define what it means for a k-round run S to satisfy an extended formula φ (written $S \models \varphi$).

1. $S \models p$, where p is a primitive proposition, if $S = (\gamma, T, \text{sent}, \text{received})$ and p is true under the truth assignment γ.
2. $S \models \sim\varphi$ if $S \not\models \varphi$.
3. $S \models \varphi_1 \wedge \varphi_2$ if $S \models \varphi_1$ and $S \models \varphi_2$.
4. $S \models K_i\varphi$ if $S' \models \varphi$ whenever $S' \sim_i S$.
5. $S \models I\varphi$ if $S' \models \varphi$ for each S' such that $S' \sim_i S$ for every $i \in \mathcal{P}$.

Intuitively, part (4) of the definition says that player i *knows* φ in a k-round run if φ is satisfied by every k-round run that player i thinks is possible. Part (5) of the definition says that φ is *implicit knowledge* in a k-round run if φ is satisfied by every k-round run that everyone thinks is possible (cf. [HM2]). It is useful to note for later use that the following formulas are valid (satisfied by every run): $K_i\varphi \Rightarrow I\varphi$ ("Anything known by player i is implicit knowledge"), and $K_i\varphi_1 \wedge K_i(\varphi_1 \Rightarrow \varphi_2) \Rightarrow K_i\varphi_2$ ("What player i knows is closed under modus ponens").

If $0 \le k' \le k$, then the k'-round prefix of a k-round run $(\gamma, T, \text{sent}, \text{received})$ is defined in the obvious way: the k'-round prefix is $(\gamma, T, \text{sent}', \text{received}')$, where $\text{sent}'(i,j,r) = \text{sent}(i,j,r)$ for each i,j and each $r \le k'$, and similarly for received$'$.

Finally, we define what it means for a message $\varphi \in \text{sent}(i,j,s)$, where $1 \le s \le k+1$, to be *i-honest* in a $(k+1)$-round run $S = (\gamma, T, \text{sent}, \text{received})$ (written $S \models_i \varphi$).

1. $S \models_i$ "I knew θ just after round r" if $S' \models K_i\theta$, where S' is the r-round prefix of S.
2. $S \models_i$ "I sent message φ to player j in round r" if $\varphi \in \text{sent}(i,j,r)$.
3. $S \models_i$ "I sent precisely the set Φ of messages to player j in round r" if $\Phi = \text{sent}(i,j,r)$.
4. $S \models_i$ "I received message φ from player j in round r" if $\varphi \in \text{received}(i,j,r)$.
5. $S \models_i$ "I received precisely the set Φ of messages from player j in round r" if $\Phi = \text{received}(i,j,r)$.
6. $S \models_i$ "Every message that I sent in round r to player j I still know to be true" if $S \models_i \varphi$ whenever $\varphi \in \text{sent}(i,j,r)$.
7. $S \models_i \sim\varphi$ if $S \not\models_i \varphi$.
8. $S \models_i \varphi_1 \wedge \varphi_2$ if $S \models_i \varphi_1$ and $S \models_i \varphi_2$.
9. $S \models_i$ "If ψ holds in round r, then in round r, the only messages I will send to player j are in the set Φ", if either (a) $r > k+1$, (b) $S' \not\models_i \psi$, where S' is the r-round prefix of S, or (c) $\text{sent}(i,j,r) \subseteq \Phi$.

The reader should note that in part (9), we are defining what it means for a message φ about the future, *which may have been sent in an early round of S*, to be honest in S. Intuitively, clause (a) of part (9) has the effect that a message about the future directly involving, say, round 17, is always considered honest before round 17.

The following lemma, whose straightforward proof is omitted, will be used later.

Lemma 2.1. *Assume that $S \sim_i S'$. Then $S \models_i \varphi$ iff $S' \models_i \varphi$, for every message φ.*

3. Conservation of implicit knowledge

In this section, we give a fundamental principle of communication-based knowledge, which says that no amount of communication in a closed system can change the implicit knowledge about nature in the system. This principle is quite robust, and holds independent of our assumptions that communication is synchronous, that communication is unreliable, that if a message is received, then it is received in the round it was sent, etc.

If φ is a formula, S is a k-round run, and $r \le k$, then we say that φ *is implicit knowledge after r rounds of S* if $S' \models I\varphi$, where S' is the r-round prefix of S.

The conservation principle for implicit knowledge: *Let φ be a propositional formula, and let S be a k-round run. Assume that $0 \leq r \leq k$ and $0 \leq s \leq k$. Then φ is implicit knowledge after r rounds of S if and only if φ is implicit knowledge after s rounds of S.*

Thus, implicit knowledge about nature never changes after the 0th round. The conservation principle is false if φ is not required to be a propositional formula, that is, if φ is not a formula about nature. For example, if φ is $K_1 p$, where p is a primitive proposition, and if player 1 learns nothing from nature (in round 0) but learns that p is true in round 1 because of a message from player 2, then φ is implicit knowledge after round 1 (it is even known by player 1 after round 1), but it is not implicit knowledge after round 0 (it is even false after round 0).

We now prove the conservation principle. It suffices to show that φ is implicit knowledge after r rounds of S if and only if φ is implicit knowledge after the 0th round of S. If φ is implicit knowledge after the 0th round of S, then it is easy to see that φ is implicit knowledge after r rounds of S (this is because nature never changes, and information about nature is never lost). Assume now that φ is not implicit knowledge after the 0th round of S. We shall show that φ is not implicit knowledge after r rounds of S. Let S be $(\gamma, T, \mathrm{sent}, \mathrm{received})$. Since φ is not implicit knowledge after the 0th round of S, it follows from our definition of satisfaction that there is some primitive state β such that (a) β does not satisfy φ, and (b) $\beta \in T(i)$ for every player i. Let S' be $(\beta, T, \mathrm{sent}, \mathrm{received})$. Thus, S' is just like S, except that the primitive state in S' is β instead of γ. It is straightforward to see that S' is a k-round run. The only nontrivial issue is to show that every message in S' is honest. But this follows from Lemma 2.1, since every message in S is honest, and S' is i-equivalent to S for every player i. Now S' does not satisfy φ, since β does not satisfy φ. Therefore, since S' is i-equivalent to S for every player i, it follows that φ is not implicit knowledge after r rounds of S. This was to be shown.

4. A new axiom

In this section, we present an interesting new axiom, which we shall show is sound. Like the conservation principle, this axiom is quite robust under a number of changes in our assumptions.

Define a *primitive state formula* to be a formula that completely describes a primitive state. For example, if there are exactly two primitive propositions, namely p and q, then up to equivalence, there are exactly four primitive state formulas, namely $p \wedge q$, $p \wedge \sim q$, $\sim p \wedge q$, and $\sim p \wedge \sim q$. The new axiom is:

$I \sim \alpha \Rightarrow (K_1 \sim \alpha \vee \ldots \vee K_n \sim \alpha),$

where α is a primitive state formula, and where $1, \ldots, n$ are all the players. Note that $K_j \sim \alpha$ appears within this new axiom for every player j. This axiom says that if it is implicit knowledge that a primitive state is impossible, then the stronger fact is true that some player knows that the primitive state is impossible. In other words, if by putting all of their information together the players could rule out the primitive state α, then some player, by himself, could have ruled out α. This is quite surprising, since we might imagine that it could happen that the reason it is implicit knowledge that a primitive state is impossible is because of some complicated combination of "high depth knowledge" of the various players (we shall define the depth of formulas shortly).

To prove the soundness of this axiom, let us consider the contrapositive. The contrapositive says that if all of the players individually think that the primitive state α is possible, then $\sim\alpha$ is not implicit knowledge. We now show that this is true about an arbitrary k-round run S. Assume that in S, all of the players think that the primitive state α is possible. So, all of the players think that α is possible after the 0th round of S. We now show that $\sim\alpha$ is not implicit knowledge after the 0th round. Let

S' be the 0-round prefix of S. Let S'' be a 0-round run in which α is the primitive state, and which is i-equivalent to S' for every player i. Thus, each player learns the same information from nature in (the 0th round of) S'' as in S'. Since all of the players think that α is possible after the 0th round of S, it is easy to see that S'' is indeed a run. Since $S'' \models \alpha$, and since $S'' \sim_i S'$ for every player i, it follows that $\sim\alpha$ is not implicit knowledge after the 0th round of S, which was to be shown. Hence, by the conservation principle for implicit knowledge, it follows that $\sim\alpha$ is not implicit knowledge after the kth round. Hence, $\sim\alpha$ is not implicit knowledge in S, which was to be shown.

Example 4.1. We now show that the formula that results by allowing α in our new axiom be a primitive proposition p, rather than a primitive state formula, is not sound if there are at least two primitive propositions. Assume that there are two primitive propositions p and q, and two players, Alice and Bob. Consider the 0-round run where both p and q are false, and where, in the 0th round, Alice learns that p and q are either both true or both false and Bob learns that q is false (but he learns nothing about p). Then $\sim p$ is implicit knowledge, since Alice and Bob could combine their information and learn that p is false. However, neither Alice nor Bob know that p is false. Thus, the new axiom does not hold if we were to let α be p. ■

We just showed that one generalization of our new axiom is not sound. We now give a sound generalization. Let us say that player i is *indifferent* to the primitive proposition p if for each truth assignment t that player i thinks is possible, he also thinks that the truth assignment t' is possible, where t' is the same as t except that p is true in t if and only if p is false in t'. Let α be a *partial state formula*, that is, a formula which describes a truth assignment to a subset X of the primitive propositions (if X were the set of all primitive propositions, then we would have a primitive state formula). If every player is indifferent to every primitive proposition that is not in X, it is not hard to show that $I\sim\alpha \Rightarrow (K_1\sim\alpha \vee ... \vee K_n\sim\alpha)$, is still sound, even though α is only a partial state formula, and not a (full) primitive state formula. This may be important in practice, where there may be infinitely many primitive propositions, but where, except for those in a small set X, every player may be indifferent to all of them.

To help understand the new axiom, we now give a general principle of implicit knowledge which has the new axiom as a corollary. We begin with some definitions. If Σ is a set of extended formulas, and σ is a single extended formula, then we say that Σ *implies* σ, written $\Sigma \models \sigma$, if every run that satisfies every member of Σ also satisfies σ. Thus, $\Sigma \models \sigma$ if there is no "counterexample" run that satisfies every member of Σ but does not satisfy σ. We may write $\Sigma \not\models \sigma$ if it is not the case that $\Sigma \models \sigma$. If Σ is a singleton $\{\tau\}$, then we may write $\tau \models \sigma$ for $\{\tau\} \models \sigma$.

The *depth* of a formula φ, denoted depth(φ), is defined as follows:

1. $\text{depth}(p) = 0$ if p is a primitive proposition
2. $\text{depth}(\sim\varphi) = \text{depth}(\varphi)$
3. $\text{depth}(\varphi_1 \wedge \varphi_2) = \max(\text{depth}(\varphi_1), \text{depth}(\varphi_2))$
4. $\text{depth}(K_i\varphi) = 1 + \text{depth}(\varphi)$

Note that we are only defining the depth for formulas, not for extended formulas.

As before, let the players be $1, ..., n$. Let us say that an extended formula φ *follows from the depth k knowledge of the players in run S* if there are formulas φ_1, ..., φ_n, each of depth at most k, such that $S \models K_i\varphi_i$ for each player i, and $\{\varphi_1, ..., \varphi_n\} \models \varphi$.

The next theorem helps give us some insight about implicit knowledge in runs. We shall show that our new axiom follows easily from it.

Theorem 4.2. *Let φ be a depth k formula that is implicit knowledge in run S, that is, $S \models I\varphi$. Then φ follows from the depth k knowledge of the players in run S.*

This theorem follows easily from a result in Section 11. It is surprising for the same reasons we gave earlier that our new axiom is surprising: we might imagine that it could happen that the reason that φ is implicit knowledge is because of some complicated combination of high depth knowledge of the various players.

We now show that our new axiom follows directly from a special case of this theorem. We need the following simple lemma.

Lemma 4.3. *Let Σ be a set of propositional formulas, and let α be a primitive state formula. If $\Sigma \models \sim\alpha$, then $\sigma \models \sim\alpha$ for some $\sigma \in \Sigma$.*

Proof. Assume that $\sigma \not\models \sim\alpha$ for each $\sigma \in \Sigma$. It follows easily that the truth assignment represented by α makes σ true for every $\sigma \in \Sigma$. It again follows easily that $\Sigma \not\models \sim\alpha$. The lemma follows. ■

We now show that the new axiom follows from Theorem 4.2 and Lemma 4.3. Assume that $I{\sim}\alpha$ holds in run S. We must show that $S \models K_i{\sim}\alpha$ for some player i. Since $\sim\alpha$ is a propositional formula, it follows from Theorem 4.2 that $\sim\alpha$ follows from the depth 0 knowledge of the players in run S. That is, there are propositional formulas φ_1, ..., φ_n such that $S \models K_i\varphi_i$ for each player i, and $\{\varphi_1, ..., \varphi_n\} \models \varphi$. By Lemma 4.3, $\varphi_i \models \sim\alpha$ for some player i. Hence, $S \models K_i{\sim}\alpha$. This was to be shown.

5. An S5-consistent formula that is not satisfiable under communication

Assume that there is only one primitive proposition p, and only two players, Alice and Bob. Let φ be the formula

(2) $K_{Alice}((p \wedge \sim K_{Alice}p \wedge \sim K_{Bob}p) \vee (\sim p \wedge \sim K_{Alice}{\sim}p \wedge K_{Bob}{\sim}p))$

This is formula (1) from the introduction, where we have replaced players 1 and 2 by Alice and Bob. It is easy to verify that φ is S5-consistent (in the sense that there is a model of S5 which satisfies this formula). In this section, we show that no run satisfies φ. Thus, $\sim\varphi$ is valid in our system. In particular, this shows that S5 is not a complete axiomatization for knowledge under message exchange. In the next section, we give two sound and complete axiomatizations (one using implicit knowledge and one not using implicit knowledge).

Let φ_1 be the formula $p \wedge \sim K_{Alice}p \wedge \sim K_{Bob}p$, which says that p is true, and that neither Alice nor Bob knows that p is true. Let φ_2 be the formula $\sim p \wedge \sim K_{Alice}{\sim}p \wedge K_{Bob}{\sim}p$, which says that p is false, that Alice does not know that p is false, and that Bob knows that p is false. Then the formula φ that we wish to show is not satisfied by any run is $K_{Alice}(\varphi_1 \vee \varphi_2)$.

It is instructive to give two proofs that φ is not satisfiable. The first proof, which is somewhat informal, proceeds as follows.

Let S be a k-round run that satisfies φ. Since everything Alice knows is true, it follows that S satisfies $\varphi_1 \vee \varphi_2$. Therefore, Alice does not know whether p is true or false in S. Also, Alice can reason to herself as follows:

> *I know that either φ_1 or φ_2 holds. Assume that φ_1 holds. Then p would be true, and Bob would not know that p is true, just as I do not know that p is true. In particular, since we would both think that it is possible that p is false, we would not have implicit knowledge that p is true. This follows immediately from the axiom $Ip \Rightarrow (K_{Alice}p \vee K_{Bob}p)$ of Section 4, where the primitive state formula α is $\sim p$. In the next round, Bob could correctly send me a message saying that he does not know that p is false. This*

would tell me that φ_2 is false, since φ_2 implies that Bob knows that p is false. Since I already know that either φ_1 or φ_2 holds, I could then deduce that φ_1 holds. But φ_1 implies that p is true, and so I would deduce that p is true. In particular, after the next round, Bob and I would have implicit knowledge that p is true. This violates the conservation principle for implicit knowledge, since I already observed that p was not implicit knowledge. This contradiction shows that φ_1 is impossible. Therefore, φ_2 holds, and so p is false. I have just proven that p is false, and so I know that p is false in run S. But this contradicts the fact that I do not know whether p is true or false in run S!

The second proof shows directly, without appealing to the notion of implicit knowledge, that φ is unsatisfiable. Let S be a k-round run that satisfies φ. Now φ implies $\sim K_{Alice}\sim\varphi_1$, since if Alice knows that φ_1 is false, then she knows that φ_2 is true, although it is clear that the formula $K_{Alice}\varphi_2$ is inconsistent. We have shown that Alice thinks that φ_1 is possible. This means that there is some k-round run S_1 that is Alice-equivalent to S and that satisfies φ_1. In particular, p is true in S_1. Let S_2 be just like S_1, except that p is false in S_2. As in the proof of the conservation principle for implicit knowledge, it is easy to see that S_2 is a k-round run, which is both Alice-equivalent and Bob-equivalent to S_1. Now $S_1 \models \varphi_1$, and so $S_1 \models p$. Hence, $S_1 \models \sim K_{Bob}\sim p$. Therefore, since S_1 and S_2 are Bob-equivalent, it follows that $S_2 \models \sim K_{Bob}\sim p$. Now S_2 is Alice-equivalent to S, since S_2 is Alice-equivalent to S_1, which is Alice-equivalent to S. So, since $S \models \varphi$, it follows that $S_2 \models \varphi_1 \vee \varphi_2$. Clearly $S_2 \not\models \varphi_1$, since $S_2 \models \sim p$. Hence, $S_2 \models \varphi_2$. In particular, $S_2 \models K_{Bob}\sim p$. But we showed that $S_2 \models \sim K_{Bob}\sim p$. This is a contradiction.

We have just given two proofs that there is no run that satisfies φ when (a) there is exactly one primitive proposition p and (b) Alice and Bob are the only players. However, if either (a) or (b) is false, then there is a run that satisfies φ. We now exhibit a run that satisfies φ if (b) is false, and leave as an amusing exercise for the reader to find a run that satisfies φ if (a) is false.

Example 5.1. Assume that there are three players (Alice, Bob, and Charlie), and one primitive proposition p. We now exhibit a 1-round run that satisfies φ. In fact, it is convenient to exhibit two such runs, S_1 and S_2. In S_1, the primitive proposition p is true; in the 0th round of run S_1, neither Alice nor Bob learn that p is true, but Charlie learns that p is true. In round 1 of S_1, Alice receives a message from Charlie saying: "In the next round, I will not send any messages to Bob" (it is easy to see that this can be viewed as one of our messages about the future). In round 2 of S_1, Alice receives a message from Bob saying: "I do not know that p is true", and a message from Charlie saying: "If p is false, then Bob knows that p is false". In S_2, the primitive proposition p is false; in the 0th round of run S_2, Alice does not learn that p is false, but both Bob and Charlie learn that p is false. In round 1 of S_2, Charlie receives a message from Bob saying: "I know that p is false". In round 1 of S_1, Alice receives a message from Charlie saying: "In the next round, I will not send any messages to Bob". In round 2 of S_1, Alice receives a message from Bob saying: "I do not know that p is true", and a message from Charlie saying: "If p is false, then Bob knows that p is false". No other messages are sent in either run. Note that the two runs are Alice-equivalent. In both runs, Alice still does not know whether p is true or false after the second round. This is because the two runs are Alice-equivalent, and in one run p is true, while in the other, p is false.

We now show that φ is satisfied by, say, run S_1. In S_1 (that is, at the end of round 2 of S_1), Alice knows that Bob does not know that p is true. This is because (a) Bob told her in round 2 that he (Bob) does not know that p is true, and (b) she knows that he did not learn that p is true in round 2, since she knows that no one sent him any messages in round 2 (Charlie promised Alice that he would not send Bob any messages, and she also knows that she certainly didn't send any messages).

Now Alice knows that there are two possibilities: (i) p is true, and (ii) p is false. In case (i), Alice knows (as we just saw) that Bob does not know that p is true, and she of course knows that she does not know that p is true; hence, Alice knows that if case (i) holds, then formula φ_1 holds. Alice also knows that if case (ii) holds, then Bob knows that p is false (since Charlie told her this, and since once Bob knows that p is false, Bob always knows that p is false.) In case (ii), she of course also knows that she does not know that p is false. Hence, Alice knows that if case (ii) holds, then formula φ_2 holds. So Alice knows that either φ_1 or φ_2 holds. Hence, run S_1 satisfies $K_{Alice}(\varphi_1 \vee \varphi_2)$; that is, run S_1 satisfies φ. ■

6. Complete axiomatizations and decision problems

In this section we give a sound and complete axiomatization for the extended formulas that are valid in runs, that is, which are satisfied by every run. We also give a sound and complete axiomatization where only formulas, rather than extended formulas, are allowed in the axioms.

We begin by presenting the classical axiom system S5 (or actually, its generalization to multiple players). Following Halpern and Moses [HM2], we refer to the system as S5$_n$ when there are n players $1, ..., n$. The axioms are:

- All substitution instances of propositional tautologies.
- $K_i\varphi \Rightarrow \varphi$ ("Whatever player i knows is true").
- $K_i\varphi \Rightarrow K_iK_i\varphi$ ("Player i knows what he knows").
- $\sim K_i\varphi \Rightarrow K_i \sim K_i\varphi$ ("Player i knows what he does not know").
- $K_i\varphi_1 \wedge K_i(\varphi_1 \Rightarrow \varphi_2) \Rightarrow K_i\varphi_2$ ("What player i knows is closed under modus ponens").

There are two rules of inference: modus ponens ("from φ_1 and $\varphi_1 \Rightarrow \varphi_2$ infer φ_2") and knowledge generalization ("from φ infer $K_i\varphi$").

We now give Halpern and Moses' system S5I_n, which they show is a sound and complete axiomatization for knowledge and implicit knowledge in Kripke structures [HM2]. S5I_n contains all of the axioms and rules of S5$_n$, along with some axioms for implicit knowledge. The first implicit knowledge axiom is:

- $K_i\varphi \Rightarrow I\varphi$ ("Whatever each individual player knows is implicit knowledge").

The remaining axioms of S5I_n say that implicit knowledge behaves like individual knowledge.

- $I\varphi \Rightarrow \varphi$
- $I\varphi \Rightarrow II\varphi$
- $\sim I\varphi \Rightarrow I \sim I\varphi$
- $I\varphi_1 \wedge I(\varphi_1 \Rightarrow \varphi_2) \Rightarrow I\varphi_2$

Let ML$_n$ (where ML stands for "Message Logic") be S5I_n, along with our new axiom from Section 4:

- $I \sim \alpha \Rightarrow (K_1 \sim \alpha \vee ... \vee K_n \sim \alpha)$,

where α is a primitive state formula. Note that $K_j \sim \alpha$ appears within this new axiom for every player j.

We are now ready to state our completeness theorem for extended formulas. Of course, since we are interested in communication, we only consider the case where there are at least two players. (We note that for the case of exactly one player, we would just add another axiom which says $I\varphi \Rightarrow K_1\varphi$, which the one player is player 1.)

Theorem 6.1. *ML$_n$ is a sound and complete axiomatization for valid extended formulas in runs with $n \geq 2$ players.*

Theorem 6.1 is hard to prove. Details will appear in a later version [FHV2] of this paper.

It is natural to ask for a sound and complete axiomatization when only formulas, rather than extended formulas, are allowed in the axioms (that is, where the axioms do not mention implicit knowledge). We now give such an axiom system, which we call $\mathrm{ML}_n{}^-$ (where the superscripted minus sign refers to the fact that implicit knowledge is not part of the language). The system consists of $S5_n$, along with a new axiom, which we shall give shortly.

Define a *pure knowledge formula* to be Boolean combination of formulas of the form $K_i\varphi$, where φ is arbitrary. For example, $K_2p \vee (K_1{\sim}K_3p \wedge {\sim}K_2{\sim}p)$ is a pure knowledge formula, but $p \wedge {\sim}K_ip$ is not. Assume that there are n players $1,\dots,n$. Our new axiom is:

(3) $K_i(\varphi \Rightarrow {\sim}\alpha) \Rightarrow K_i(\varphi \Rightarrow (K_1{\sim}\alpha \vee \dots \vee K_n{\sim}\alpha))$,

for all players i, all pure knowledge formulas φ, and all primitive state formulas α.

Note that $K_j{\sim}\alpha$ appears within this new axiom for every player j. A loose translation of this axiom is "If player i knows that some 'pure knowledge' φ is incompatible with some primitive state α, then player i knows the stronger fact that the pure knowledge φ forces some player to know that the primitive state α is impossible". We now show that this somewhat unintuitive axiom (3) is sound. If not, then let S be a run that does not satisfy (3). So $S \models K_i(\varphi \Rightarrow {\sim}\alpha)$ and $S \not\models K_i(\varphi \Rightarrow \psi)$, where ψ is the formula $K_1{\sim}\alpha \vee \dots \vee K_n{\sim}\alpha$. Since $S \not\models K_i(\varphi \Rightarrow \psi)$, there is a run S' such that $S \sim_i S'$ where $S' \models \varphi$ and $S' \models {\sim}\psi$. Since $I{\sim}\alpha \Rightarrow \psi$ is valid (this is our new Message Logic axiom), it follows that $S' \not\models I{\sim}\alpha$. Therefore, there is a run S'' such that $S'' \sim_j S'$ for every player j, where $S'' \models \alpha$. Since $S'' \sim_j S'$ for every player j, it is straightforward to show that every pure knowledge formula satisfied by S' is satisfied by S''. Therefore, $S'' \models \varphi$. By transitivity of $\sim_i$, we know that $S'' \sim_i S$. Therefore, since $S \models K_i(\varphi \Rightarrow {\sim}\alpha)$, it follows that $S'' \models \varphi \Rightarrow {\sim}\alpha$. But we already showed that $S'' \models \varphi$ and $S'' \models \alpha$. This is a contradiction. Hence, (3) is sound.

The formula

(4) $K_{Alice}(({\sim}K_{Bob}p \wedge {\sim}K_{Bob}{\sim}p) \Rightarrow p) \Rightarrow K_{Alice}(({\sim}K_{Bob}p \wedge {\sim}K_{Bob}{\sim}p) \Rightarrow (K_{Alice}p \vee K_{Bob}p))$.

is an instance of the new axiom (3). It is straightforward to verify that (4) implies that formula (2) in Section 5 is unsatisfiable.

As before, if we allow α in the new axiom (3) to be a primitive proposition, rather than a primitive state formula, then the axiom does not remain sound, even if α is the only primitive proposition that appears in φ.

Theorem 6.2. $\mathrm{ML}_n{}^-$ *is a sound and complete axiomatization for valid formulas in runs.*

Theorem 6.2, like Theorem 6.1, is hard to prove. Details will appear in a later version [FHV2] of this paper.

It is interesting to consider what would happen if we were to allow only messages about the past (that is, if in our definition of the class of messages, we were to eliminate clause 2, which defines messages about the future).

Theorem 6.3. *If only messages about the past are allowed, and if there are exactly two players, then our axiomatizations are still complete (that is, Theorem 6.1 and Theorem 6.2 still hold). However, our axiomatizations are not complete if there are at least three players.*

Proof. The first part of the theorem holds, since our completeness proof in the case of two players does not require messages about the future. We now give an example that shows that if we were to restrict messages to be about the past, then our axiomatizations are not complete for at least three players. Assume that there are three players, say, Alice, Bob, and Charlie. Let p be a primitive proposition. If every message is about the past, then it is impossible to arrive at a situation where Alice knows that Bob knows p and that Charlie does not know p: this is because Alice never knows whether Bob has just told Charlie that p is true. It is instructive to see how, by allowing messages about the future, it is possible for Alice to know that Bob knows p and that Charlie does not know p (for pedagogical reasons, we shall use slightly more general "messages about the future" than we have already defined, although of course this is not essential). Bob sends a message to Alice, telling her that he knows p and that he has never sent a message to Charlie and never will; Charlie sends a message to Alice telling her that he does not know p; and Alice does not send any messages to Charlie. ■

We close this section by giving the complexity of the decision problems for ML_n and ML_n^-. It is known [HM2] that the decision problem for $S5_n$ (with $n \geq 2$) is PSPACE-complete. The next theorem says that ML_n and ML_n^- are no harder (and no easier) than $S5_n$. Of course, in ML_n and ML_n^-, we are interested only in the case when $n \geq 2$, that is, when there are at least two players.

Theorem 6.4. *The decision problems for* ML_n *and* ML_n^- *are PSPACE-complete (when* $n \geq 2$*).*

7. What if communication is reliable?

In this section, we briefly consider the situation where communication is reliable, that is, where messages can never be never lost. In this case, the set of states of knowledge that can arise is greatly restricted, as we shall show. Nevertheless, the conservation principle and our axioms are still sound, as the reader can verify.

Let us say that two runs are *equivalent* if they satisfy the same extended formulas.

Theorem 7.1. *Assume that there are only two players, and that communication is reliable. Then every run is equivalent to a 1-round run where both players send exactly one message.*

Proof. Assume that the two players are players 1 and 2. Let $S = (\gamma, T, \mathrm{sent}, \mathrm{received})$ be a run (where communication is reliable). Of course, received is now redundant, since $\mathrm{received}(i,j,r) = \mathrm{sent}(j,i,r)$ for every i,j,r. If V is a set of primitive states, then let us say that V is *possible initial information about nature for player 1* if there is some primitive state γ' such that $S' = (\gamma', T', \mathrm{sent}, \mathrm{received})$ is a run, where $T'(1) = V$ and $T'(2) = T(2)$. Intuitively, "V is possible initial information about nature for player 1" if as far as player 2 is concerned, all of player 1's messages would have been legal if V would have been player 1's initial information about nature. Similarly, we define what if means for a set V of primitive states to be *possible initial information about nature for player 2.*

If V is a set of primitive states, then let τ_V be the formula

(5) $(\bigwedge\{K_1{\sim}\alpha\colon \alpha \notin V\}) \wedge (\bigwedge\{{\sim}K_1{\sim}\alpha\colon \alpha \in V\})$.

Intuitively, τ_V says that player 1 thinks that precisely the primitive states in V are possible. Let φ_1 be the formula

(6) $\bigvee\{\tau_V\colon V \text{ is possible initial information about nature for player } 1\}$

Intuitively, φ_1 gives precisely the information that player 2 has about player 1 in S. Similarly, define φ_2. Let $S^{\#} = (\gamma, T, \mathrm{sent}', \mathrm{received}')$ be a 1-round run (with γ and T the same as in S), where the only message that player 1 sends is "I knew φ_1 just after round 0", and similarly for player 2. It is not hard to see that $S^{\#}$ is a 1-round run that fulfills the conditions of the theorem. ■

Corollary 7.2. *Assume that there are only two players, and that communication is reliable. Let the set of primitive states be fixed. Then there are only a finite number of distinct equivalence classes of runs (where two runs are in the same equivalence class if they satisfy the same extended formulas).*

Proof. Since there are only a finite number of distinct 1-round runs of the type $S^{\#}$ as defined in the proof of Theorem 7.1, the result follows immediately. ■

Theorem 7.1 and Corollary 7.2 contrast with the situation in which communication is unreliable. For, let S_k be a k-round run in which Alice tells Bob in the first round that Alice knows that the primitive proposition p is true, where Bob acknowledges to Alice in the second round that he received this message from Alice, where Alice acknowledges to Bob in the third round that she received Bob's acknowledgment, and so on through the kth round. It is easy to see that S_k is not equivalent to any $(k-1)$-round run, and that no two of the runs S_k are equivalent.

We note that it follows from Corollary 7.2 that our axiomatizations in Section 6 are not complete (although, as we have noted, they are sound). In particular, it can be shown that if there are exactly two players and if α is a primitive state formula, then the formula $(K_1K_2\alpha \wedge K_2K_1\alpha) \Rightarrow K_1K_2K_1\alpha$ is valid.

The reader should note that the results of this section apply only to the case of exactly two players. The case of three or more players is currently under investigation.

8. Changing the class of messages

In Theorem 6.3 and Section 7, we considered some effects of changing the class of messages. In this section, we briefly discuss the effect more generally. Most importantly, it turns out that our class of messages is rich enough that increasing the class in a reasonable way does not cause the set of axioms to change. We now discuss what we mean by this claim.

If all of the assumptions we have made hold, except that the class of messages is changed to $\mathcal{M}'$, then let $R(\mathcal{M}')$ be the set of all runs (involving these messages), and let $A(\mathcal{M}')$ be the resulting complete axiomatization for the valid extended formulas. Let $\mathcal{M}$ be the class of messages we have allowed in this paper, and assume that $\mathcal{M} \subseteq \mathcal{M}'$. Assume further that our axioms remain sound when $\mathcal{M}'$ is the class of messages (that is, assume that $A(\mathcal{M}) \subseteq A(\mathcal{M}')$. It turns out that the completeness proof then shows that $A(\mathcal{M}) = A(\mathcal{M}')$, that is, our axiomatization is still complete.

It is instructive to give a false "proof" of this fact. It is easy to convince oneself that if $\mathcal{M}_1 \subseteq \mathcal{M}_2$, then $A(\mathcal{M}_2) \subseteq A(\mathcal{M}_1)$. After all, if we have more possible runs, then there should be fewer valid formulas. Therefore, in our case, $A(\mathcal{M}') \subseteq A(\mathcal{M})$. Since by assumption $A(\mathcal{M}) \subseteq A(\mathcal{M}')$, it follows that $A(\mathcal{M}) = A(\mathcal{M}')$, as desired.

It is indeed true that if the class of models increases, then the set of axioms can only decrease (or stay the same). However, in our case, a "model" is *not* a run, but rather, a pair $(S, \mathscr{R})$, where S is a run and $\mathscr{R}$ is a set of runs. Namely, $\mathscr{R}$ is the set of runs that are conceivable; for us, $\mathscr{R}$ is $R(\mathcal{M}')$, where $\mathcal{M}'$ is the class of messages. So the set of models is not necessarily comparable, even if $\mathcal{M}_1 \subseteq \mathcal{M}_2$.

Example 8.1. Let p and q be primitive propositions. Let $\mathcal{M}_1$ contain exactly one message, namely, "I know p", and let $\mathcal{M}_2$ contain exactly two messages, namely, "I know p" and "I know q". Let σ be the formula $\sim K_1K_2q$, and let τ be the formula $\sim((K_1K_2p) \wedge K_1(q \wedge \sim K_3K_2q))$. It is easy to see that σ is in $A(\mathcal{M}_1)$ (since there are no messages involving q), but not in $A(\mathcal{M}_2)$. We now sketch a proof that τ is in $A(\mathcal{M}_2)$ but not in $A(\mathcal{M}_1)$. We first show that τ is in $A(\mathcal{M}_2)$. If not, then let S be a run in $R(\mathcal{M}_2)$ that satisfies $(K_1K_2p) \wedge K_1(q \wedge \sim K_3K_2q)$. Since S satisfies K_1K_2p, there has been at least one round of message exchange in S. Since player 1 knows q, for all player 1 knows, the following occurred: player 2 learned that q was true in the 0th round, and told player 3 in the next round that he (player 2) knows q. In this case, player 3 would know that player 2 knows q. Since as far as player 1 is concerned, this gives a possible run, it follows that player 1 does not know that player 3 does not know that player 2 knows q. This is a contradiction. We now show that τ is not in $A(\mathcal{M}_1)$. For, let S be a run where player 1 learns in round 0 that q is true, and where player 2 tells player 1 that player 2 knows that p is true. Then S satisfies $\sim\tau$, since player 1 knows that player 3 cannot know that player 2 knows q (since there are no messages involving q). Thus, $A(\mathcal{M}_1)$ and $A(\mathcal{M}_2)$ are incomparable, even though $\mathcal{M}_1 \subseteq \mathcal{M}_2$. ∎

9. Knowledge structures and knowledge worlds

In this section we briefly review the definition of *knowledge worlds* from [FHV1]. We first discuss them informally.

Example 9.1. Assume there are two players, Alice and Bob, and that there is only one primitive proposition p. There are various "levels" of knowledge. At the "0th level" ("nature"), assume that p is true. The 1st level tells each player's knowledge about nature. For example, Alice's knowledge at the 1st level could be "I (Alice) don't know whether p is true or false", and Bob's could be "I (Bob) know that p is true". The 2nd level tells each player's knowledge about the other player's knowledge about nature. For example, Alice's knowledge at the 2nd level could be "I know that Bob knows whether p is true or false", and Bob's could be "I don't know whether Alice knows p". Thus, Alice knows that either p is true and Bob knows it, or else p is false and Bob knows it. At the 3rd level, Alice's knowledge could be "I know that Bob does not know whether I know about p". This can continue for arbitrarily many levels. ∎

We now give the formal definition of a (knowledge) world. We assume a fixed finite set of primitive propositions, and a fixed finite set $\mathcal{P}$ of players. A *0th-order knowledge assignment*, f_0, is a truth assignment to the primitive propositions. We call $<f_0>$ a *1-ary world* (since its "length" is 1). Assume inductively that k-ary worlds $<f_0,...,f_{k-1}>$ have been defined. Let W_k be the set of all k-ary worlds. A *kth-order knowledge assignment* is a function f_k: $\mathcal{P} \rightarrow 2^{W_k}$. Intuitively, f_k associates with each player a set of "possible k-ary worlds". There are certain semantic restrictions on f_k, which we shall list shortly. These restrictions enforce the properties of knowledge mentioned above. We call $<f_0,...,f_k>$ a *$(k+1)$-ary world.* (Although we shall deal only with worlds, we note for completeness that an infinite sequence $<f_0 f_1 f_2,...>$ is called a *knowledge structure* if each prefix $<f_0,...,f_{k-1}>$ is a k-ary world for each k.)

Before we list the restrictions on f_k, let us reconsider Example 9.1. In that example, f_0 is the truth assignment that makes p true. Also, $f_1(\text{Alice}) = \{p,\bar{p}\}$ (where by p (respectively, $\bar{p}$) we mean the 1-ary world $<f_0>$, where f_0 is the truth assignment that makes p true (respectively, false)), and $f_1(\text{Bob}) = \{p\}$. Saying $f_1(\text{Alice}) = \{p,\bar{p}\}$ means that Alice does not know whether p is true or false. We can write the 2-ary world $<f_0 f_1>$ as $<p, (\text{Alice} \mapsto \{p,\bar{p}\}, \text{Bob} \mapsto \{p\})>$. Let us denote this 2-ary world by w_1. Let w_2 be the 2-ary world $<\bar{p}, (\text{Alice} \mapsto \{p,\bar{p}\}, \text{Bob} \mapsto \{\bar{p}\})>$, and let w_3 be $<p, (\text{Alice} \mapsto \{p\}, \text{Bob} \mapsto \{p\})>$. In Example 9.1, $f_2(\text{Alice}) = \{w_1, w_2\}$, since Alice thinks both w_1 (where p is true and Bob knows it) and w_2 (where p is false and Bob knows it) are possible worlds. Similarly, $f_2(\text{Bob}) = \{w_1, w_3\}$, since Bob thinks both w_1 (where p is true and Alice does not know it) and w_3 (where p is true and Alice knows it) are possible worlds.

A $(k+1)$-ary world $<f_0,...,f_k>$ must satisfy the following restrictions for each player i:

(K1) $<f_0, ..., f_{k-1}> \in f_k(i)$, if $k \geq 1$ ("The real k-ary world is one of the possibilities, for each player"). In our example, we see that indeed $p \in f_1$(Alice) and $p \in f_1$(Bob). Furthermore, $w_1 \in f_2$(Alice) and $w_1 \in f_2$(Bob), where we recall that w_1 is the "real" 2-ary world $<f_0, f_1>$.

(K2) If $<g_0, ..., g_{k-1}> \in f_k(i)$, and $k > 1$, then $g_{k-1}(i) = f_{k-1}(i)$ ("Player i knows exactly what he knows"). Let us consider our example. Alice thinks there are two possible 2-ary worlds, namely w_1 and w_2, since $f_2(\text{Alice}) = \{w_1, w_2\}$. If we write w_2 as $<g_0, g_1>$, then indeed g_1(Alice) $= \{p, \bar{p}\} = f_1$(Alice), as required. Intuitively, although Alice has doubts about Bob's knowledge, she has no doubts about her own knowledge. Thus, in all 2-ary worlds she considers possible, her knowledge is identical, namely, she does not know whether p is true or false.

(K3) $<g_0, ..., g_{k-2}> \in f_{k-1}(i)$ iff there is a $(k-1)$st-order knowledge assignment g_{k-1} such that $<g_0, ..., g_{k-2}, g_{k-1}> \in f_k(i)$, if $k > 1$ ("i's higher-order knowledge is an extension of i's lower-order knowledge"). In our example, since Alice thinks either p or $\bar{p}$ is possible, there is some 2-ary world she thinks possible (namely, w_1) in which p is true, and there is some 2-ary world she thinks possible (namely, w_2) in which p is false. Conversely, because she thinks w_1 and w_2 are both possible, it follows that she thinks either p or $\bar{p}$ is possible.

We now define what it means for an $(r+1)$-ary world $<f_0, ..., f_r>$ to satisfy formula φ, written $<f_0, ..., f_r> \models \varphi$, if $r \geq \text{depth}(\varphi)$, where depth($\varphi$) is defined as in Section 4.

1. $<f_0, ..., f_r> \models p$, where p is a primitive proposition, if p is true under the truth assignment f_0.
2. $<f_0, ..., f_r> \models \sim\varphi$ if $<f_0, ..., f_r> \not\models \varphi$.
3. $<f_0, ..., f_r> \models \varphi_1 \wedge \varphi_2$ if $<f_0, ..., f_r> \models \varphi_1$ and $<f_0, ..., f_r> \models \varphi_2$.
4. $<f_0, ..., f_r> \models K_i\varphi$ if $<g_0, ..., g_{r-1}> \models \varphi$ for each $<g_0, ..., g_{r-1}> \in f_r(i)$.

Let us again consider Example 9.1. Let w_1 and w_2 be, as before, the two 2-ary worlds that Alice considers possible. Then $w_1 \models K_{\text{Bob}}p$, since according to w_1, the only 1-ary world Bob considers possible is $<p>$. Similarly, $w_2 \models K_{\text{Bob}}\sim p$. Hence, both w_1 and w_2 satisfy $(K_{\text{Bob}}p \vee K_{\text{Bob}}\sim p)$. Since both of the 2-ary worlds that Alice considers possible satisfy $(K_{\text{Bob}}p \vee K_{\text{Bob}}\sim p)$, it follows that in our example $<f_0, f_1, f_2> \models K_{\text{Alice}}(K_{\text{Bob}}p \vee K_{\text{Bob}}\sim p)$.

The following crucial lemma shows a certain robustness in the definition of the satisfaction of a formula in a world.

Lemma 9.2. *Assume that* depth(φ) $= k$ *and* $r \geq k$. *Then* $<f_0, ..., f_r> \models \varphi$ *iff* $<f_0, ..., f_k> \models \varphi$.

10. Message-based knowledge worlds

In this section we define a restricted class of knowledge worlds, which we call "message-based knowledge worlds". The reason for the name is that these turn out to be precisely the worlds that arise under message exchange.

First, 1-ary (respectively, 2-ary) message-based knowledge worlds are exactly the same as 1-ary (respectively, 2-ary) knowledge worlds, as defined in Section 9. Then $(k+1)$-ary message-based knowledge worlds (for $k \geq 2$) are $(k+1)$-ary knowledge worlds $<f_0, ..., f_k>$ that satisfy the following additional restrictions for each player i: (a) $f_k(i)$ is a set of k-ary message-based knowledge worlds, and (b) whenever $<g_0, g_1, ..., g_{k-1}> \in f_k(i)$, and whenever $g_0' \in g_1(j)$ for every player j, then $<g_0', g_1, ..., g_{k-1}> \in f_k(i)$. What condition (b) says is that whenever player i thinks that a world $w = <g_0, g_1, ..., g_{k-1}>$ is possible, then he also thinks that every world $<g_0', g_1, ..., g_{k-1}>$ is also possible, for every truth assignment g_0' consistent with everyone's knowledge about nature in w, that is, consistent with g_1. Intuitively, a good way to think of this is that instead of players "imagining" possible worlds that look like $<g_0,g_1,...,g_{k-1}>$, every player imagines "worlds" $<g_1,...,g_{k-1}>$, where automatically for every truth assignment g_0 consistent with g_1, the player thinks that the world $<g_0,g_1,...,g_{k-1}>$ is possible.

The next theorem shows that message-based knowledge worlds correspond to the knowledge gained in runs.

Theorem 10.1. *For each k-round run S and each nonnegative integer r, there is an r-ary message-based knowledge world* $w = <f_0,f_1,...,f_{r-1}>$ *such that S and w satisfy precisely the same formulas of depth* $r-1$ *or less. Conversely, for each r-ary message-based knowledge world w, there is a k-round run S (for some k) such that S and w satisfy precisely the same formulas of depth* $r-1$ *or less.*

The difficult step of the proof consists of taking an arbitrary message-based knowledge world and producing a run, including a complete description of what messages each player sends in each round and which messages are lost, such that the world and the run satisfy the same formulas (of appropriate depth).

11. Implicit knowledge in message-based knowledge worlds

Implicit knowledge of a group of players is the knowledge that can be obtained by pooling together the group's knowledge. Let $<f_0,...f_k>$ be a $(k+1)$-ary world. For each player i, the set $f_k(i)$ consists of all the k-ary worlds that player i thinks are possible. Thus, implicitly the players think that precisely the k-ary worlds in $\bigcap_{i\in\mathcal{P}} f_k(i)$ are the possible ones. If φ is a formula of depth r, where $r \leq k$, then we say that $<f_0,...f_k>$ satisfies $I\varphi$ if φ is satisfied by all the k-ary worlds in $\bigcap_{i\in\mathcal{P}} f_k(i)$.

Consider now an extension $<f_0,...f_k,f_{k+1}>$ of $<f_0,...f_k>$. In view of Lemma 9.2, we might be tempted to believe that $<f_0,...f_k,f_{k+1}>$ satisfies $I\varphi$ if and only if $<f_0,...f_k>$ satisfies $I\varphi$. Unfortunately, this is not the case; instead, implication holds in only one direction. Thus, if $<f_0,...f_k>$ satisfies $I\varphi$, then also $<f_0,...f_k,f_{k+1}>$ satisfies $I\varphi$. But it is possible that $<f_0,...f_k>$ does not satisfy $I\varphi$, while $<f_0,...f_k,f_{k+1}>$ satisfies $I\varphi$. This can happen because a k-ary world w can be a member of $\bigcap_{i\in\mathcal{P}} f_k(i)$, even though no extension of it is a member of $\bigcap_{i\in\mathcal{P}} f_{k+1}(i)$. (Note, however, that if a world is in $f_k(i)$, then some extension of it is in $f_{k+1}(i)$, by restriction (K3) on knowledge worlds.)

Put differently, the extended formula $I\varphi$, where φ is a formula of depth k, is not a formula of depth $k+1$, but rather it is a formula of arbitrary depth. To understand this, recall our example of implicit knowledge from the introduction. If Alice knows ψ and Bob knows $\psi \Rightarrow \varphi$, then together they have implicit knowledge of φ, though neither of them might individually know φ. Now even though the formula φ is of depth k, the formula ψ can be of arbitrary depth, so the implicit knowledge of φ is essentially knowledge of arbitrary depth. Unfortunately, the framework of knowledge structures and knowledge worlds requires that formulas be assigned a well-defined depth, so this framework cannot handle implicit knowledge (in particular Lemma 9.2 would fail for extended formulas if we were to define $\text{depth}(I\varphi) = 1 + \text{depth}(\varphi)$).

Surprisingly, in the context of message-based knowledge, implicit knowledge is quite "well behaved". The basis for this is the following property of message-based knowledge worlds.

Lemma 11.1. *In a message-based knowledge world, if* $k > 1$, *then* $<g_0,...,g_{k-2}> \in \bigcap_{i\in\mathcal{P}} f_{k-1}(i)$ *if and only if there is* g_{k-1} *such that* $<g_0,...,g_{k-2},g_{k-1}> \in \bigcap_{i\in\mathcal{P}} f_k(i)$.

Note the strong similarity between Lemma 11.1 and restriction (K3) on knowledge worlds. As we noted earlier, Theorem 4.2 follows from Lemma 11.1.

As we said before, Lemma 11.1 does not hold for arbitrary knowledge worlds. For message-based communication, it follows from Lemma 11.1 that if φ is a formula of depth k, then we can define $I\varphi$ to be of depth $k+1$. We can then define the semantics of extended formulas in message-based knowledge structures in a straightforward way, and extend Theorem 10.1 to deal with extended formulas. Details will be given in a later version of this paper.

12. Concluding remarks

The main point of the paper is that we cannot reason about knowledge without taking into account how the knowledge is acquired in the first place. We have focused here on distributed systems where knowledge is acquired via unreliable message exchange. Certain knowledge states were shown to be unattainable in this model. We have characterized the attainable knowledge states and axiomatized the formulas that are valid in in such states. It turns out that the basic feature of message-based knowledge is conservation of implicit knowledge. Thus our results, as well as recent results in [DM, DM, PR, RP], indicate that implicit knowledge is a fundamental concept to the understanding and analysis of distributed systems.

In this paper we have focused on a particular model of communication. Our main assumptions are as follows.

1. Nature never changes.
2. Communication is synchronous, and proceeds in rounds.
3. Communication is unreliable.
4. If a message is received at all, then it is received in the round it was sent.
5. Messages are taken from a particular class of messages.
6. Players "receive information about nature", and then all information is obtained by communication, without any further input "from the outside".

Though these assumptions are quite natural, one may want to consider other models of communication, where the above assumptions are changed. We believe that the issue of how communication affects knowledge deserves a great deal of further study. Thus beyond its technical contributions, this paper opens up an interesting, and we hope fruitful, line of research.

13. Acknowledgments

The authors are grateful to Cynthia Dwork, Yoram Moses, and Larry Stockmeyer for helpful suggestions. We are especially grateful to Joe Halpern for making major simplifications to our completeness proofs, and for pointing out that increasing the class of messages might, in the general case, destroy completeness.

14. Bibliography

[Au] R. J. Aumann, Agreeing to disagree, *Annals of Statistics* 4,6 (1976), pp. 1236-1239.

[CM] M. Chandy and J. Misra, How processes learn, *Proc. 4th ACM Symp. on Principles of Distributed Computing*, 1985, pp. 204-214.

[Dw] C. Dwork, Bounds on fundamental problems in parallel and distributed computation, Ph.D. thesis, Cornell University, 1984.

[DM] C. Dwork and Y. Moses, Knowledge and common knowledge in a Byzantine environment I: crash failures, this proceedings, 1986.

[DFIL] C. Dwork, M.J. Fischer, N. Immerman, and N.A. Lynch, A theory of protocols, unpublished notes, 1984.

[FHV1] R. Fagin, J. Y. Halpern, and M. Y. Vardi, A model-theoretic analysis of knowledge, *Proc. 25th IEEE Symp. on Foundations of Computer Science*, West Palm Beach, Florida, 1984, pp. 268-278.

[FHV2] R. Fagin, J. Y. Halpern, and M. Y. Vardi, To appear.

[FV] R. Fagin and M. Y. Vardi, An internal semantics for modal logic, *Proc. 17th ACM Symp. on Theory of Computing*, 1985, pp. 305-315.

[HF] J. Y. Halpern and R. Fagin, A formal model of knowledge, action, and communication in distributed systems: preliminary report, *Proc. ACM Symp. on Principles of Distributed Computation*, 1985, pp. 224-236.

[HM1] J. Y. Halpern and Y.O. Moses, Knowledge and common knowledge in a distributed environment, *Proc. 3rd ACM Symp. on Principles of Distributed Computing*, 1984, pp. 50-61.

[HM2] J. Y. Halpern and Y.O. Moses, A guide to the modal logics of knowledge and belief, *Proc. International Joint Conference on Artificial Intelligence (IJCAI-85)*, 1985, pp. 480-490.

[Hi] J. Hintikka, *Knowledge and belief*. Cornell University Press, 1962.

[Kr] S. Kripke, Semantical analysis of modal logic, *Zeitschrift fur Mathematische Logik und Grundlagen der Mathematik* 9 (1963), pp. 67-96.

[Leh] D. Lehman, Knowledge, common knowledge, and related puzzles, *Proc. 3rd ACM Symp. on Principles of Distributed Computing*, 1984, pp. 467-480.

[MH] J. McCarthy and P. Hayes, Some philosophical problems from the standpoint of artificial intelligence, in *Machine Intelligence 4*, (ed. D. Michie), American Elsevier, 1969, pp. 463-502.

[Pa] R. Parikh, Logics of knowledge, games, and nonmonotonic logic, *Proc. FST-TCS*, 1984, Lecture Notes in Computer Science - vol. 181, pp. 202-222.

[PR] R. Parikh and R. Ramanujam, Distributed processing and the logic of knowledge, *Proc. Workshop on Logics of Programs*, Brooklyn, 1985.

[RP] S. Rosenschein and F. Pereira, Knowledge and action in situated automata, unpublished manuscript, 1985.

[Sa] L. J. Savage, *The Foundations of Statistics*, Wiley, New York, 1954.