Common Knowledge Revisited*

Ronald Fagin
IBM Almaden Research Center
650 Harry Road
San Jose, CA 95120–6099
`fagin@almaden.ibm.com`

Joseph Y. Halpern†
IBM Almaden Research Center
650 Harry Road
San Jose, CA 95120–6099
`halpern@almaden.ibm.com`

Yoram Moses‡
Dept. of Applied Math. and CS
The Weizmann Institute of Science
76100 Rehovot, Israel
`yoram@cs.weizmann.ac.il`

Moshe Y. Vardi§
Dept. of Computer Science
Rice University
Houston, TX 77005-1892
`vardi@cs.rice.edu`

Abstract: We consider the common-knowledge paradox raised in [HM90]: common knowledge is necessary for coordination, but common knowledge is unattainable in the real world because of temporal imprecision. We discuss two solutions to this paradox: (1) modeling the world with a coarser granularity, and (2) relaxing the requirements for coordination.

1 Introduction

The notion of *common knowledge,* where everyone knows, everyone knows that everyone knows, etc., has proven to be fundamental in various disciplines, including Philosophy [Lew69], Artificial Intelligence [MSHI79], Game Theory [Aum76], Psychology [CM81], and Distributed Systems [HM90]. This key notion was first studied by the philosopher David Lewis [Lew69] in the context of conventions. Lewis pointed out that in order for something to be a convention, it must in fact be common knowledge among the members of a group. (For example, the

*This material is based on our recently published book [FHMV95].

†Work supported in part by the Air Force Office of Scientific Research (AFSC), under Contract F49620-91-C-0080.

‡Part of this research was performed while this author was on sabbatical at Oxford. His work is supported in part by a Helen and Milton A. Kimmelman career development chair.

§This research was done while this author was at the IBM Almaden Research Center. Author's URL: `http://www.cs.rice.edu/~vardi` convention that green means “go” and red means “stop” is presumably common knowledge among the drivers in our society.)

Common knowledge also arises in discourse understanding [CM81]. Suppose Ann asks Bob “What did you think of the movie?” referring to a showing of *Monkey Business* they have just seen. Not only must Ann and Bob both know that “the movie” refers to *Monkey Business*, but Ann must know that Bob knows (so that she can be sure that Bob will give a reasonable answer to her question), Bob must know that Ann knows that Bob knows (so that Bob knows that Ann will respond appropriately to his answer), and so on. In fact, by a closer analysis of this situation, it can be shown that there must be common knowledge of what movie is meant in order for Bob to answer the question appropriately.

Finally, as shown in [HM90], common knowledge also turns out to be a prerequisite for agreement and coordinated action. This is precisely what makes it such a crucial notion in the analysis of interacting groups of agents. On the other hand, in practical settings common knowledge is impossible to achieve. This puts us in a somewhat paradoxical situation, in that we claim both that common knowledge is a prerequisite for agreement and coordinated action and that it cannot be attained. We discuss two answers to this paradox: (1) modeling the world with a coarser granularity, and (2) relaxing the requirements for coordination.

2 Two puzzles

We start by discussing two well-known puzzles that involve attaining common knowledge. The first is the “muddy children” puzzle (which goes back at least to [GS58], although the version we consider here is taken from [Bar81]).

The story goes as follows: Imagine n children playing together. Some, say k of them, get mud on their foreheads. Each can see the mud on others but not on his own forehead. Along comes the father, who says, “At least one of you has mud on your forehead,” thus expressing a fact known to each of them before he spoke (if $k > 1$). The father then asks the following question, over and over: “Does any of you know whether you have mud on your own forehead?” Assuming that all the children are perceptive, intelligent, truthful, and that they answer simultaneously, what will happen?

There is a straightforward proof by induction that the first $k - 1$ times he asks the question, they will all say “No,” but then the k^{th} time the children with muddy foreheads will all answer “Yes.” Let us denote the fact “at least one child has a muddy forehead” by p. Notice that if $k > 1$, i.e., more than one child has a muddy forehead, then every child can see at least one muddy forehead, and the children initially all know p. Thus, it would seem that the father does not provide the children with any new information, and so he should not need to tell them that p holds when $k > 1$. But this is false! What the father provides is common knowledge. If exactly k children have muddy foreheads, then it is straightforward to see that $E^{k-1}p$ holds before the father speaks, but $E^k p$ does not (here $E^k\varphi$ means φ, if $k = 0$, and everyone knows $E^{k-1}\varphi$, if $k \geq 1$). The father’s statement actually converts the children’s state of knowledge from $E^{k-1}p$ to Cp (here Cp means that there is common knowledge of p). With this extra knowledge, they can deduce whether their foreheads are muddy.

In the muddy children puzzle, the children do not actually need common knowledge; $E^k p$ suffices for them to figure out whether they have mud on their foreheads. On the other hand, the *coordinated attack* problem introduced by Gray [Gra78] provides an example where common knowledge is truly necessary. In this problem, two generals, each commanding a division of an army, want to attack a common enemy. They will win the battle only if they attack the enemy simultaneously; if only one division attacks, it will be defeated. Thus, the generals want to coordinate their attack. Unfortunately, the only way they have of communicating is by means of messengers, who might get lost or captured by the enemy.

Suppose a messenger sent by General A reaches General B with a message saying "*attack at dawn*." Should General B attack? Although the message was in fact delivered, General A has no way of knowing that it was delivered. A must therefore consider it possible that B did not receive the message (in which case B would definitely not attack). Hence A will not attack given his current state of knowledge. Knowing this, and not willing to risk attacking alone, B cannot attack based solely on receiving A's message. Of course, B can try to improve matters by sending the messenger back to A with an acknowledgment. Even if the messenger reaches A, similar reasoning shows that neither A nor B will attack at this point either. In fact, Yemini and Cohen [YC79] proved, by induction on the number of messages, that no number of successful deliveries of acknowledgments to acknowledgments can allow the generals to attack. Halpern and Moses [HM90] showed the relationship between coordinated attack and common knowledge, and used this to give a "knowledge-based" proof of Yemini and Cohen's result. Specifically, assume that the generals behave according to some predetermined deterministic protocol; that is, a general's actions (what messages he sends and whether he attacks) are a deterministic function of his history and the time on his clock. Assume further that in the absence of any successful communication, neither general will attack. Halpern and Moses then prove the following theorem:

Theorem 2.1 [HM90] *A correct protocol for the coordinated attack problem must have the property that whenever the generals attack, it is common knowledge that they are attacking.*

Halpern and Moses then define the notion of a system where *communication is not guaranteed*. Roughly speaking, this means (1) it is always possible that from some point on, no messages will be received, and (2) if a processor (or general) i does not get any information to the contrary (by receiving some message), then i considers it possible that none of its messages were received. In particular, in the coordinated attack problem as stated, communication is not guaranteed. Halpern and Moses then prove that in such a system, nothing can become common knowledge unless it is also common knowledge in the absence of communication. This implies the impossibility of coordinated attack:

Theorem 2.2 [HM90] *Any correct protocol for the coordinated attack problem guarantees that neither general ever attacks.*

Common knowledge of φ is defined to be the infinite conjunction of the formulas $E^k\varphi$. This definition seems to suggest that common knowledge has an "inherently infinite" nature. Indeed, for a fact that is not common knowledge to become common knowledge, each participating agent must come to know an infinite collection of new facts. Could this be one of the reasons that common knowledge is impossible to attain in this case? Not really.

In practice, there is always a finite bound on the number of possible local states of an agent in a real-world system. A *finite-state system* is one where each agent's set of possible local states is finite. Fischer and Immerman [FI86] showed that in a finite-state system, common knowledge is equivalent to E^k for a sufficiently large k. Nevertheless, the result that common knowledge is not attainable if communication is not guaranteed applies equally well to finite-state systems (as do our later results on the unattainability of common knowledge). Thus, in such cases, $E^k\varphi$ is unattainable for some sufficiently large k. (Intuitively, k is large enough so that the agents cannot count up to k; that is, k is tantamount to infinity for these agents.) So the unattainability of common knowledge in this case is not due to the fact that common knowledge is defined in terms of an infinite conjunction.

3 Common Knowledge and Uncertainty

As we have seen, common knowledge cannot be attained when communication is not guaranteed. Halpern and Moses show further that common knowledge cannot be attained in a system in which communication *is* guaranteed, but where there is no bound on the time it takes for messages to be delivered. It would seem that when all messages are guaranteed to be delivered within a fixed amount of time, say one second, attaining common knowledge should be a simple matter. But things are not always as simple as they seem; even in this case, uncertainty causes major difficulties.

Consider the following example: Assume that two agents, Alice and Bob, communicate over a channel in which (it is common knowledge that) message delivery is guaranteed. Moreover, suppose that there is only slight uncertainty concerning message delivery times. It is commonly known that any message sent from Alice to Bob reaches Bob within ε time units. Now suppose that at some point Alice sends Bob a message μ that does not specify the sending time in any way. Bob does not know initially that Alice sent him a message. We assume that when Bob receives Alice's message, he knows that it is from her. How do Alice and Bob's state of knowledge change with time?

Let *sent*(μ) be the statement that Alice sent the message μ. After ε time units, we have $K_A K_B$*sent*(μ), that is, Alice knows that Bob knows that she sent the message μ. And clearly, this state of knowledge does not occur before ε time units. Define $(K_A K_B)^k$*sent*(μ) by letting it be *sent*(μ) for $k = 0$, and $K_A K_B (K_A K_B)^{k-1}$*sent*(μ) for $k \geq 1$. It is not hard to verify that $(K_A K_B)^k$*sent*(μ) holds after $k\varepsilon$ time units, and does not hold before then. In particular, common knowledge of *sent*(μ) is never attained. This may not seem too striking when we think of ε that is relatively large, say a day, or an hour. The argument, however, is independent of the magnitude of ε, and remains true even for small values of ε. Even if Alice and Bob are guaranteed that Alice's message arrives within one nanosecond, they still never attain common knowledge that her message was sent!

Now let us consider what happens if both Alice and Bob use the *same* clock, and suppose that, instead of sending μ, Alice sends at time m a message μ' that specifies the sending time, such as

"This message is being sent at time m; μ."

Recall that it is common knowledge that every message sent by Alice is received by Bob within ε time units. When Bob receives μ', he knows that μ' was sent at time m. Moreover, Bob's receipt of μ' is guaranteed to happen no later than time $m + \varepsilon$. Since Alice and Bob use the same clock, it is common knowledge at time $m + \varepsilon$ that it is $m + \varepsilon$. It is also common knowledge that any message sent at time m is received by time $m + \varepsilon$. Thus, at time $m + \varepsilon$, the fact that Alice sent μ' to Bob is common knowledge.

Note that in the first example common knowledge will never hold regardless of whether ε is a day, an hour, or a nanosecond. The slight uncertainty about the sending time and the message transmission time prevents common knowledge of μ from ever being attained in this scenario. What makes the second example so dramatically different? When a fact φ is common knowledge, everybody must know that it is. It is impossible for agent i to know that φ is common knowledge without agent j knowing it as well. This means that the transition from φ not being common knowledge to its being common knowledge must involve a *simultaneous* change in all relevant agents' knowledge. In the first example, the uncertainty makes such a simultaneous transition impossible, while in the second, having the same clock makes a simultaneous transition possible and this transition occurs at time $m + \varepsilon$. These two examples help illustrate the connection between simultaneity and common knowledge and the effect this can have on the attainability of common knowledge. We now formalize and further explore this connection.

4 Simultaneous Events

The Alice and Bob examples illustrate how the transition from a situation in which a fact is not common knowledge to one where it is common knowledge requires simultaneous events to take place at all sites of the system. The relationship between simultaneity and common knowledge, however, is even more fundamental than that. We saw by example earlier that actions that must be performed simultaneously by all parties, such as attacking in the coordinated attack problem, become common knowledge as soon as they are performed: common knowledge is a prerequisite for simultaneous actions. In this section, we give a result that says that a fact's becoming common knowledge requires the occurrence of simultaneous events at different sites of the system. Moreover, the results say that in a certain technical sense, the occurrence of simultaneous events is necessarily common knowledge. This will demonstrate the strong link between common knowledge and simultaneous events.

To make this claim precise, we need to formalize the notion of simultaneous events. We begin by briefly reviewing the framework of [FHMV95] for modeling multi-agent systems. We assume that at each point in time, each agent is in some *local state*. Informally, this local state encodes the information available to the agent at this point. In addition, there is an *environment* state, that keeps track of everything relevant to the system not recorded in the agents' states.

A *global state* is an $(n+1)$-tuple $(s_e, s_1, \ldots, s_n)$ consisting of the environment state s_e and the local state s_i of each agent i. A *run* of the system is a function from time (which, for ease of exposition, we assume ranges over the natural numbers) to global states. Thus, if r is a run, then $r(0), r(1), \ldots$ is a sequence of global states that, roughly speaking, is a complete description of how the system evolves over time in one possible execution of the system. We take a *system* to consist of a set of runs. Intuitively, these runs describe all the possible sequences of events that could occur in a system.

Given a system $\mathcal{R}$, we refer to a pair (r, m) consisting of a run $r \in \mathcal{R}$ and a time m as a *point*. If $r(m) = (s_e, s_1, \ldots, s_n)$, we define $r_i(m) = s_i$, for $i = 1, \ldots, n$; thus, $r_i(m)$ is process i's local state at the point (r, m). We say two points (r, m) and (r', m') are *indistinguishable* to agent i, and write $(r, m) \sim_i (r', m')$, if $r_i(m) = r'_i(m')$, i.e., if agent i has the same local state at both points. Finally, we define an *interpreted system* to be a pair $(\mathcal{R}, \pi)$ consisting of a system $\mathcal{R}$ together with a mapping π that associates a truth assignment to the primitive propositions with each global state.

An interpreted system can be viewed as a Kripke structure: the points are the possible worlds, and $\sim_i$ plays the role of the accessibility relation. We give semantics to knowledge formulas in interpreted systems just as in Kripke structures: Given a point (r, m) in an interpreted system $\mathcal{I} = (\mathcal{R}, \pi)$, we have $(\mathcal{I}, r, m) \models K_i\varphi$ if $(\mathcal{I}, r', m') \models \varphi$ for all points (r', m') such that $(r', m') \sim_i (r, m)$. Notice that under this interpretation, an agent knows φ if φ is true at all the situations the system could be in, given the agent's current information (as encoded by his local state). Since $\sim_i$ is an equivalence relation, knowledge in this framework satisfies the S5 axioms. If G is a set of agents, we define E_G ("everyone in the group G knows") by saying $(\mathcal{I}, r, m) \models E_G\varphi$ if $(\mathcal{I}, r, m) \models K_i\varphi$ for every $i \in G$. We define C_G ("it is common knowledge among the agents in G") by saying $(\mathcal{I}, r, m) \models C_G\varphi$ if $(\mathcal{I}, r, m) \models (E_G)^k\varphi$ for every k. When G is the set of all agents, we may write E for E_G, and C for C_G. We write $\mathcal{I} \models \varphi$ if $(\mathcal{I}, r, m) \models \varphi$ for every point (r, m) of the system $\mathcal{I}$.

We now give a few more definitions, all relative to a fixed interpreted system $\mathcal{I} = (\mathcal{R}, \pi)$. Let S denote the set of points of the system $\mathcal{R}$. Define an *event* in $\mathcal{R}$ to be a subset of S; intuitively, these are the points where the event e holds. An event e is said to *hold* at a point (r, m) if $(r, m) \in e$. Of special interest are events whose occurrence is reflected in an agent's local state. More formally, an event e is *local to* i (in interpreted system $\mathcal{I}$) if there is a set L_i^e of i's local states such that for all points (r, m) we have $(r, m) \in e$ iff $r_i(m) \in L_i^e$. The events of sending a message, receiving a message, and performing an internal action are examples of local events for agent i. We remark that the definition of a local event does not imply that an event that is local to i cannot also be local to j. In order to be local to both agents, it only needs to be reflected in the local states of both agents.

Certain events depend only on the global state. An event e is a *state event* if there is a set $\mathcal{G}^e$ of global states such that for all points (r, m) we have $(r, m) \in e$ iff $r(m) \in \mathcal{G}^e$. It is easy to see that local events are state events. More generally, a state event is one that depends only on what is recorded in the local states of the agents and the state of the environment. We associate with every state event e a primitive proposition ψ_e that is true at the global state $r(m)$ if and only if $(r, m) \in e$. This is well-defined, because it follows easily from the definition of state events that if e is a state event and (r, m) and (r', m') are points where $r(m) = r'(m')$, then $(r, m) \in e$ if and only if $(r', m') \in e$.

We can similarly associate with every formula φ an event $\mathtt{ev}_{\mathcal{I}}(\varphi) = \{(r, m) \,|\, (\mathcal{I}, r, m) \models \varphi\}$. The event $\mathtt{ev}_{\mathcal{I}}(\varphi)$ thus holds exactly when φ holds. We call $\mathtt{ev}_{\mathcal{I}}(\varphi)$ *the event of* φ *holding (in* $\mathcal{I}$*)*. It is easy to check that an event e is local to i if and only if $K_i\psi_e$ holds, that is, if and only if i knows that e is holding. Moreover, the event of $K_i\varphi$ holding is always a local event for i.

We are now ready to address the issue of simultaneous events. Intuitively, two events are simultaneous if they occur at the same points. Our interest in simultaneity is primarily in the context of coordination. Namely, we are interested in events that are local to different agents and are coordinated in time. Thus, we concentrate on events whose occurrence is simultaneously reflected in the local state of the agents. More formally, we define an *event ensemble for* G (or just *ensemble* for short) to be a mapping $\mathbf{e} : i \mapsto e_i$, assigning to every agent $i \in G$ an event e_i local to i. An ensemble $\mathbf{e}$ for G is said to be *perfectly coordinated* if the local events in $\mathbf{e}$ hold simultaneously; formally, if $(r, m) \in \mathbf{e}(i)$ for some $i \in G$, then $(r, m) \in \mathbf{e}(j)$ for all $j \in G$. Thus, the ensemble $\mathbf{e}$ for G is perfectly coordinated precisely if $\mathbf{e}(i) = \mathbf{e}(j)$ for all $i, j \in G$. Since an event e_i that is local to agent i is defined in terms of a set $L_i^{e_i}$ of states local to agent i, the ensemble $\mathbf{e}$ for G is perfectly coordinated if all the agents in G enter their respective sets $L_i^{e_i}$ simultaneously. Thus, the events in a perfectly coordinated ensemble are simultaneous.

An example of a perfectly coordinated ensemble is the set of local events that correspond to the ticking of a global clock, if the ticking is guaranteed to be reflected simultaneously at all sites of a system. Another example is the event of shaking hands: being a mutual action, the handshakes of the parties are perfectly coordinated.

Given an ensemble $\mathbf{e}$ for G, the proposition $\psi_{\mathbf{e}(i)}$ corresponds to the state event $\mathbf{e}(i)$ holding. We also define $\psi_{\mathbf{e}} = \bigvee_{i \in G} \psi_{\mathbf{e}(i)}$. Thus, $\psi_{\mathbf{e}}$ is true whenever one of the state events $\mathbf{e}(i)$ holds.

Proposition 4.1 *Let* $\mathcal{I}$ *be an interpreted system and* G *a set of agents.*

(a) For every formula φ*, the ensemble* $\mathbf{e}$ *for* G *defined by* $\mathbf{e}(i) = \mathtt{ev}_{\mathcal{I}}(K_i C_G \varphi)$ *is perfectly coordinated.*

(b) If $\mathbf{e}$ *is a perfectly coordinated ensemble for* G*, then* $\mathcal{I} \models \psi_{\mathbf{e}} \Rightarrow C_G \psi_{\mathbf{e}}$.

(In fact, $K_i C_G \varphi$ in part (a) of Proposition 4.1 is logically equivalent to $C_G \varphi$, but we write $K_i C_G \varphi$ for analogy with a later proposition.) Proposition 4.1 precisely captures the close correspondence between common knowledge and simultaneous events. It asserts that the local events that correspond to common knowledge are perfectly coordinated, and the local events in a perfectly coordinated ensemble are common knowledge when they hold. Notice that part (a) implies in particular that the transitions from $\neg K_i C_G \varphi$ to $K_i C_G \varphi$, for $i \in G$, must be simultaneous. Among other things, this helps clarify the difference between the two examples considered in Section 3: In the first example, Alice and Bob cannot attain common knowledge of *sent*(μ) because they are unable to make such a simultaneous transition, while in the second example they can (and do).

The close relationship between common knowledge and simultaneous actions is what makes common knowledge such a useful tool for analyzing tasks involving coordination and agreement. It also gives us some insight into how common knowledge arises. For example, the fact that a public announcement has been made is common knowledge, since the announcement is heard simultaneously by everyone. (Strictly speaking, of course, this is not quite true; we return to this issue in Section 6.) More generally, simultaneity is inherent in the notion of *copresence*. As a consequence, when people sit around a table, the existence of the table, as well as the nature of the objects on the table, are common knowledge.

Proposition 4.1 formally captures the role of simultaneous actions in making agreements and conventions common knowledge. As we discussed earlier, common knowledge is inherent in agreements and conventions. Hand shaking, face-to-face or telephone conversation, and a simultaneous signing of a contract are standard ways of reaching agreements. They all involve simultaneous actions and have the effect of making the agreement common knowledge.

5 Temporal Imprecision

As we illustrated previously and formalized in Proposition 4.1, simultaneity is inherent in the notion of common knowledge (and vice versa). It follows that simultaneity is a prerequisite for attaining common knowledge. Alice and Bob's failure to reach common knowledge in the first example above can therefore be blamed on their inability to perform a simultaneous state transition. As might be expected, the fact that simultaneity is a prerequisite for attaining common knowledge has additional consequences. For example, in many distributed systems each process possesses a clock. In practice, in any distributed system there is always some uncertainty regarding the relative synchrony of the clocks and regarding the precise message transmission times. This results in what is called the *temporal imprecision* of the system. The amount of temporal imprecision in different systems varies, but it can be argued that every practical system will have some (possibly very small) degree of imprecision. Formally, a given system $\mathcal{R}$ is said to have *temporal imprecision* if for all runs $r \in \mathcal{R}$, times m, and sets G of processes with $|G| \geq 2$, there exist processes $i, j \in G$ with $i \neq j$, a run $r' \in \mathcal{R}$, and a time m' such that $r'_i(m') = r_i(m)$ while $r'_j(m') = r_j(m+1)$. Intuitively, in a system with temporal imprecision, i is uncertain about j's clock reading; at the point (r, m), process i cannot tell whether j's clock is characterized by j's local state at (r, m) or j's local state at $(r, m+1)$. Techniques from the distributed-systems literature [DHS86, HMM85] can be used to show that any system in which, roughly speaking, there is some initial uncertainty regarding relative clock readings and uncertainty regarding exact message transmission times must have temporal imprecision.

Systems with temporal imprecision turn out to have the property that no protocol can guarantee to synchronize the processes' clocks perfectly. As we now show, events cannot be perfectly coordinated in systems with temporal imprecision either. These two facts are closely related.

We define an ensemble $\mathbf{e}$ for G in $\mathcal{I}$ to be *nontrivial* if there exist a run r in $\mathcal{I}$ and times m, m' such that $(r, m) \in \cup_{i \in G} \mathbf{e}(i)$ while $(r, m') \notin \cup_{i \in G} \mathbf{e}(i)$. Thus, if $\mathbf{e}$ is a perfectly coordinated ensemble for G, it is *trivial* if for each run r of the system and for each agent $i \in G$, the events in $\mathbf{e}(i)$ hold either at all points of r or at no point of r. The definition of systems with temporal imprecision implies the following:

Proposition 5.1 *In a system with temporal imprecision there are no nontrivial perfectly coordinated ensembles for G, if $|G| \geq 2$.*

We thus have the following corollary.

Corollary 5.2 [HM90] *Let $\mathcal{I}$ be a system with temporal imprecision, let φ be a formula, and let $|G| \geq 2$. Then for all runs r and times m we have $(\mathcal{I}, r, m) \models C_G\varphi$ iff $(\mathcal{I}, r, 0) \models C_G\varphi$.*

In simple terms, Corollary 5.2 states that no fact can become common knowledge during a run of a system with temporal imprecision. If the units by which time is measured in our model are sufficiently small, then all practical distributed systems have temporal imprecision. For example, if we work at the nanosecond level, then there is bound to be some uncertainty regarding exact message transmission times. On the other hand, if we model time at the level of minutes, this uncertainty may disappear. As a result, Corollary 5.2 implies that no fact can ever become common knowledge in practical distributed systems. Carrying this argument even further, we can view essentially all real-world scenarios as situations in which true simultaneity cannot be guaranteed. For example, the children in the muddy children puzzle neither hear nor comprehend the father simultaneously. There is bound to be some uncertainty about how long it takes each of them to process the information. Thus, according to our earlier discussion, the children in fact do not attain common knowledge of the father's statement.

We now seem to have a paradox. On the one hand, we have argued that common knowledge is unattainable in practical contexts. On the other hand, given our claim that common knowledge is a prerequisite for agreements and conventions and the observation that we do reach agreements and conventions are maintained, it seems that common knowledge *is* attained in practice.

Where is the catch? How can we explain this discrepancy between our practical experience and our technical results? In the next two sections, we consider two resolutions to this paradox. The first rests on the observation that if we model time at a sufficiently coarse level, we can and do attain common knowledge. The question then becomes when and whether it is appropriate to model time in this way. The second says that, although we indeed cannot attain common knowledge, we can attain close approximations of it, and this suffices for our purposes.

6 The Granularity of Time

Given the complexity of the real world, any mathematical model of a situation must abstract away many details. A useful model is typically one that abstracts away as much of the irrelevant detail as possible, leaving all and only the relevant aspects of a situation. When modeling a particular situation, it can often be quite difficult to decide the level of granularity at which to model time. The notion of time in a run rarely corresponds to real time. Rather, our choice of the granularity of time is motivated by convenience of modeling. Thus, in a distributed application, it may be perfectly appropriate to take a round to be sufficiently long for a process to send a message to all other processes, and perhaps do some local computation as well.

As we have observed, the argument that every practical system has some degree of temporal imprecision holds only relative to a sufficiently fine-grained model of time. For Proposition 5.1 and Corollary 5.2 to apply, time must be represented in sufficiently fine detail for temporal imprecision to be reflected in the model. If a model has a coarse notion of time, then simultaneity, and hence common knowledge, are often attainable. For example, in synchronous systems (those where the agents have access to a shared clock, so that, intuitively, the time is common knowledge) there is no temporal imprecision. As an example, consider a simplified model of the muddy children problem. The initial states of the children and the father describe what they see; later states describe everything they have heard. All communication proceeds in rounds. In round 1, if there is at least one muddy child, a message to this effect is sent to all children. In the odd-numbered rounds 1, 3, 5, . . . , the father sends to all children the message "Does any of you know whether you have mud on your own forehead?" The children respond "Yes" or "No" in the even-numbered rounds. In this simplified model, the children do attain common knowledge of the father's statement (after the first round). If, however, we "enhance" the model to take into consideration the minute details of the neural activity in the children's brains, and considered time on, say, a millisecond scale, the children would not be modeled as hearing the father simultaneously. Moreover, the children would not attain common knowledge of the father's statement. We conclude that whether a given fact becomes common knowledge at a certain point, or in fact whether it *ever* becomes common knowledge, depends in a crucial way on the model being used. While common knowledge may be attainable in a certain model of a given real world situation, it becomes unattainable once we consider a more detailed model of *the same situation*.

When are we justified in reasoning and acting as if common knowledge is attainable? This reduces to the question of when we can argue that one model—in our case a coarser or less detailed model—is "as good" as another, finer, model. The answer, of course, is "it depends on the intended application." Our approach for deciding whether a less detailed model is as good as another, finer, model, is to assume that there is some "specification" of interest, and to consider whether the finer model satisfies the same specification as the coarser model. For example, in the muddy children puzzle, our earlier model implicitly assumed that the children all hear the father's initial statement and later questions simultaneously. We can think of this as a coarse model where, indeed, the children attain common knowledge. For the fine model, suppose instead that every time the father speaks, it takes somewhere between 8 and 10 milliseconds for each child to hear and process what the father says, but the exact time may be different for each child, and may even be different for a given child every time the father speaks. Similarly, after a given child speaks, it takes between 8 and 10 milliseconds for the other children and the father to hear and process what he says. (While there is nothing particularly significant in our choice of 8 and 10 milliseconds, it is important that a child does not hear any other child's response to the father's question before he utters his own response.) The father does not ask his k^{th} question until he has received the responses from all children to his $(k-1)^{\text{st}}$ question.

The specification of interest for the muddy children puzzle is the following: A child says "Yes" if he knows whether he is muddy and says "No" otherwise. This specification is satisfied in particular when each child follows the protocol that if he sees k muddy children, then he responds "No" to the father's first k questions and "Yes" to all the questions after that. This specification is true in both the coarse model and the fine model. Therefore, we consider the coarse model adequate. If part of the specification had been that the children answer simultaneously, then the coarse model would not have been adequate. For a more formal presentation of our approach, see [FHMV95].

The observation that whether or not common knowledge is attainable depends in part on how we model time was made in a number of earlier papers [FI86, HM90, Kur86, Nei88, NT93]. Our approach formalizes this observation and offers a rigorous way to determine when the coarse model is adequate.

7 Approximations of Common Knowledge

Section 4 shows that common knowledge captures the state of knowledge resulting from simultaneous events. It also shows, however, that in the absence of events that are guaranteed to hold simultaneously, common knowledge is not attained. In Section 6, we tried to answer the question of when we can reason and act as if certain events were simultaneous. But there is another point of view we can take. There are situations where events holding at different sites need not happen simultaneously; the level of coordination required is weaker than absolute simultaneity. For example, we may want the events to hold at most a certain amount of time apart. It turns out that just as common knowledge is the state of knowledge corresponding to perfect coordination, there are states of shared knowledge corresponding to other forms of coordination. We can view these states of knowledge as approximations of true common knowledge. It is well known that common knowledge can be defined in terms of a fixed point, as well as an infinite conjunction. As shown in [HM90], $C_G\varphi$ is equivalent to $\nu x[E_G(\varphi \wedge x)]$, where νx is the *greatest fixed-point operator*.[1] As we shall see, the approximations of common knowledge have similar fixed-point definitions. Fortunately, while perfect coordination is hard to attain in practice, weaker forms of coordination are often attainable. This is one explanation as to why the unattainability of common knowledge might not spell as great a disaster as we might have originally expected. This section considers two of these weaker forms of coordination, and their corresponding states of knowledge.

Let us return to the first Alice and Bob example. Notice that if $\varepsilon = 0$, then Alice and Bob attain common knowledge of *sent*(μ) immediately after the message is sent. In this case, it is guaranteed that once the message is sent, both agents immediately know the contents of the message, as well as the fact that it has been sent. Intuitively, it seems that the closer ε is to 0,

[1] Formal definitions of this operator can be found in [FHMV95, HM90].

the closer Alice and Bob's state of knowledge should be to common knowledge. Compare the situation when $\varepsilon > 0$ with $\varepsilon = 0$. As we saw, if $\varepsilon > 0$ then Alice does not know that Bob received her message immediately after she sends the message. She does, however, know that *within ε time units* Bob will receive the message and know both the contents of the message and that the message has been sent. The sending of the message results in a situation where, within ε time units, everyone knows that the situation holds. This is analogous to the fact that common knowledge corresponds to a situation where everyone knows that the situation holds. This suggests that the state of knowledge resulting in the Alice and Bob scenario should involve a fixed point of some sort. We now formalize a notion of coordination related to the Alice and Bob example, and define an approximation of common knowledge corresponding to this type of coordination.

An ensemble $\mathbf{e}$ for G is said to be *ε-coordinated* (in a given system $\mathcal{I}$) if the local events in $\mathbf{e}$ never hold more than ε time units apart; formally, if $(r, m) \in \mathbf{e}(i)$ for some $i \in G$, then there exists an interval $I = [m', m' + \varepsilon]$ such that $m \in I$ and for all $j \in G$ there exists $m_j \in I$ for which $(r, m_j) \in \mathbf{e}(j)$. Note that ε-coordination with $\varepsilon = 0$ is perfect coordination. While it is essentially infeasible in practice to coordinate events so that they hold simultaneously at different sites of a distributed system, ε-coordination is often attainable in practice, even in systems where there is uncertainty in message delivery time. Moreover, when ε is sufficiently small, there are many applications for which ε-coordination is practically as good as perfect coordination. For example, instead of requiring a simultaneous attack in the coordinated attack problem, it may be sufficient to require only that the two divisions attack within a certain ε-time bound of each other. This is called an *ε-coordinated attack*.

More generally, ε-coordination may be practically as good as perfect coordination for many instances of agreements and conventions. One example of ε-coordination results from a message being broadcast to all members of a group G, with the guarantee that it will reach all of the members within ε time units of one another. In this case it is easy to see that when an agent receives the message, she knows the message has been broadcast, and knows that within ε time units each of the members of G will have received the message, and will know that within $\varepsilon \ldots$

Let ε be arbitrary. We say that *within an ε interval everyone in G knows φ*, denoted $E^{\varepsilon}_{G}\varphi$, if there is an interval of ε time units containing the current time such that each process comes to know φ at some point in this interval. Formally, $(\mathcal{I}, r, m) \models E^{\varepsilon}_{G}\varphi$ if there exists an interval $I = [m', m' + \varepsilon]$ such that $m \in I$ and for all $i \in G$ there exists $m_i \in I$ for which $(\mathcal{I}, r, m_i) \models K_i\varphi$. Thus, in the case of Alice and Bob, we have $\mathcal{I} \models \textit{sent}(\mu) \Rightarrow E^{\varepsilon}_{\{A,B\}}\textit{sent}(\mu)$. We define ε-common knowledge, denoted by C^{ε}_{G}, using a greatest fixed-point operator: $C^{\varepsilon}_{G}\varphi =_{\text{def}} \nu x[E^{\varepsilon}_{G}(\varphi \wedge x)]$. Notice how similar this definition is to the fixed-point definition of common knowledge. The only change is in replacing E_G by E^{ε}_{G}.

Just as common knowledge is closely related to perfect coordination, ε-common knowledge is related to ε-coordination. We now make this claim precise. The next proposition is analogous to Proposition 4.1.

Proposition 7.1 *Let $\mathcal{I}$ be an interpreted system and G a set of agents.*

(a) *For every formula φ, the ensemble* **e** *for G defined by* $\mathbf{e}(i) = \mathsf{ev}_{\mathcal{I}}(K_i C_G^{\varepsilon}\varphi)$ *is ε-coordinated.*

(b) *If* **e** *is an ε-coordinated ensemble for G, then $\mathcal{I} \models \psi_{\mathbf{e}} \Rightarrow C_G^{\varepsilon}\psi_{\mathbf{e}}$.*

Since in the coordinated attack problem message delivery is not guaranteed, it can be shown that the generals cannot achieve even ε-coordinated attack. On the other hand, if messages are guaranteed to be delivered within ε units of time, then ε-coordinated attack can be accomplished. General A simply sends General B a message saying "*attack*" and attacks immediately; General B attacks upon receipt of the message.

Although ε-common knowledge is useful for the analysis of systems where the uncertainty in message communication time is small, it is not quite as useful in the analysis of systems where message delivery may be delayed for a long period of time. In such systems, rather than perfect or ε-coordination, what can often be achieved is *eventual* coordination. An ensemble $\mathbf{e}$ for G is *eventually coordinated* (in a given system $\mathcal{I}$) if, for every run of the system, if some event in $\mathbf{e}$ holds during the run, then all events in $\mathbf{e}$ do. More formally, if $(r, m) \in \mathbf{e}(i)$ for some $i \in G$, then for all $j \in G$ there exists some m_j for which $(r, m_j) \in \mathbf{e}(j)$. An example of an eventual coordination of G consists of the delivery of (copies of) a message broadcast to every member of G in a system with message delays. An agent receiving this message knows the contents of the message, as well as the fact that each other member of G must receive the message at some point in time, either past, present, or future.

Eventual coordination gives rise to *eventual* common knowledge, denoted by $C_G^{\diamond}$, and defined by $C_G^{\diamond}\varphi =_{\text{def}} \nu x[E_G^{\diamond}(\varphi \wedge x)]$. Here we define $E_G^{\diamond}\varphi$ to hold at $(\mathcal{I}, r, m)$ if for each $i \in G$ there is some time m_i such that $(\mathcal{I}, r, m_i) \models K_i\varphi$. Thus, $E_G^{\diamond}$ can be viewed as the limit of E_G^{ε} as ε approaches infinity. It is straightforward to show that $C_G^{\diamond}$ is related to eventual coordination just as C_G is related to simultaneous events, and C_G^{ε} to ε-coordination. Interestingly, although C_G^{ε} is definable as an infinite conjunction, it can be shown that $C_G^{\diamond}$ is not [FHMV95]. We really need to use fixed points here; cf. [Bar88].

Just as ε-coordinated attack is a weakening of the simultaneity requirement of coordinated attack, a further weakening of the simultaneity requirement is given by *eventually coordinated attack*. This requirement says that if one of the two divisions attacks, then the other division eventually attacks. If messages are guaranteed to be delivered eventually, then even if there is no bound on message delivery time, an eventually coordinated attack can be carried out.

The notions of ε-common knowledge and of eventual common knowledge are from [HM90]. Our contribution here is in introducing ensembles as a formalization of the concept of coordination and in showing that approximations of common knowledge correspond to approximations of coordination. We note also that other approximations to common knowledge have been considered, including timestamped common knowledge [HM90], probabilistic common knowledge [BD87, FH94, HT93, KPN90, MS89], concurrent common knowledge [PT92], and continual common knowledge [HMW90]. All these notions can be shown to correspond to approximations of coordination.

8 Summary

The central theme of this paper is an attempt to resolve the paradox of common knowledge raised in [HM90]: Although common knowledge can be shown to be a prerequisite for day-to-day activities of coordination and agreement, it can also be shown to be unattainable in practice. The resolution of this paradox leads to a deeper understanding of the nature of common knowledge and simultaneity, and shows once again the importance of the modeling process. In particular, it brings out the importance of the granularity at which we model time, and stresses the need to consider the applications for which these notions are being used. Moreover, by using the notion of event ensembles, we are able to clarify the tight relationship between common knowledge and coordination.

References

[Aum76] R. J. Aumann. Agreeing to disagree. *Annals of Statistics*, 4(6):1236–1239, 1976.

[Bar81] J. Barwise. Scenes and other situations. *Journal of Philosophy*, 78(7):369–397, 1981.

[Bar88] J. Barwise. Three views of common knowledge. In M. Y. Vardi, editor, *Proc. Second Conference on Theoretical Aspects of Reasoning about Knowledge*, pages 365–379. Morgan Kaufmann, San Francisco, Calif., 1988.

[BD87] A. Brandenburger and E. Dekel. Common knowledge with probability 1. *Journal of Mathematical Economics*, 16:237–245, 1987.

[CM81] H. H. Clark and C. R. Marshall. Definite reference and mutual knowledge. In A. K. Joshi, B. L. Webber, and I. A. Sag, editors, *Elements of discourse understanding*. Cambridge University Press, Cambridge, U.K., 1981.

[DHS86] D. Dolev, J. Y. Halpern, and H. R. Strong. On the possibility and impossibility of achieving clock synchronization. *Journal of Computer and System Sciences*, 32(2):230–250, 1986.

[FH94] R. Fagin and J. Y. Halpern. Reasoning about knowledge and probability. *Journal of the ACM*, 41(2):340–367, 1994.

[FHMV95] R. Fagin, J. Y. Halpern, Y. Moses, and M. Y. Vardi. *Reasoning about Knowledge*. MIT Press, Cambridge, Mass., 1995.

[FI86] M. J. Fischer and N. Immerman. Foundations of knowledge for distributed systems. In J. Y. Halpern, editor, *Theoretical Aspects of Reasoning about Knowledge: Proc. 1986 Conference*, pages 171–186. Morgan Kaufmann, San Francisco, Calif., 1986.

[Gra78] J. Gray. Notes on database operating systems. In R. Bayer, R. M. Graham, and G. Seegmuller, editors, *Operating Systems: An Advanced Course*, Lecture Notes in Computer Science, Vol. 66. Springer-Verlag, Berlin/New York, 1978. Also appears as IBM Research Report RJ 2188, 1978.

[GS58] G. Gamow and M. Stern. *Puzzle Math*. Viking Press, New York, 1958.

[HM90] J. Y. Halpern and Y. Moses. Knowledge and common knowledge in a distributed environment. *Journal of the ACM*, 37(3):549–587, 1990. A preliminary version appeared in *Proc. 3rd ACM Symposium on Principles of Distributed Computing*, 1984.

[HMM85] J. Y. Halpern, N. Megiddo, and A. Munshi. Optimal precision in the presence of uncertainty. *Journal of Complexity*, 1:170–196, 1985.

[HMW90] J. Y. Halpern, Y. Moses, and O. Waarts. A characterization of eventual Byzantine agreement. In *Proc. 9th ACM Symp. on Principles of Distributed Computing*, pages 333–346, 1990.

[HT93] J. Y. Halpern and M. R. Tuttle. Knowledge, probability, and adversaries. *Journal of the ACM*, 40(4):917–962, 1993.

[KPN90] P. Krasucki, R. Parikh, and G. Ndjatou. Probabilistic knowledge and probabilistic common knowledge (preliminary report). In Z. W. Ras, M. Zemankova, and M. L. Emrich, editors, *Methodologies for Intelligent Systems*, volume 5, pages 1–8. Elsevier Science Publishing Co., Inc., The Hague, 1990.

[Kur86] R. Kurki-Suonio. Towards programming with knowledge expressions. In *Proc. 13th ACM Symp. on Principles of Programming Languages*, pages 140–149, 1986.

[Lew69] D. Lewis. *Convention, A Philosophical Study*. Harvard University Press, Cambridge, Mass., 1969.

[MS89] D. Monderer and D. Samet. Approximating common knowledge with common beliefs. *Games and Economic Behavior*, 1:170–190, 1989.

[MSHI79] J. McCarthy, M. Sato, T. Hayashi, and S. Igarishi. On the model theory of knowledge. Technical Report STAN-CS-78-657, Stanford University, 1979.

[Nei88] G. Neiger. Knowledge consistency: a useful suspension of disbelief. In M. Y. Vardi, editor, *Proc. Second Conference on Theoretical Aspects of Reasoning about Knowledge*, pages 295–308. Morgan Kaufmann, San Francisco, Calif., 1988.

[NT93] G. Neiger and S. Toueg. Simulating real-time clocks and common knowledge in distributed systems. *Journal of the ACM*, 40(2):334–367, 1993.

[PT92] P. Panangaden and S. Taylor. Concurrent common knowledge: defining agreement for asynchronous systems. *Distributed Computing*, 6(2):73–93, 1992.

[YC79] Y. Yemini and D. Cohen. Some issues in distributed processes communication. In *Proc. of the 1st International Conf. on Distributed Computing Systems*, pages 199–203, 1979.