**Logical Omniscience and Common Knowledge;
WHAT do we know and what do WE know?**

Rohit Parikh
City University of New York

http://www.sci.brooklyn.cuny.edu/cis/parikh/

Abstract: *Two difficult issues for the logic of knowledge have been logical omniscience and common knowledge. Our existing logics of knowledge based on Kripke structures seem to justify logical omniscience, but we know that in real life it does not exist. Also, common knowledge* appears *to be needed for certain real life procedures to work. But it seems quite implausible that it actually exists in real people.*

We suggest two procedure based *semantics for knowledge which seem to take care of both these issues in a relatively realistic way.*

What this suggests is that if we really want to understand knowledge, then existing customs *and* plans *must play a greater role than we are used to assigning them.*

1 Introduction

The logic of knowledge most popular these days uses Kripke structures with an accessibility relation R, typically assumed to be reflexive, symmetric, and transitive. If we are talking about belief rather than knowledge, then R would be serial, transitive, and euclidean. Then some formula ϕ is said to be known (believed) at state s iff ϕ is true at all states R-accessible from s.

While this approach has been very successful in producing rich logics of knowledge and belief, some difficulties also arise which make one doubt their relation to actual human reasoning and knowledge. Thus, for instance, it is impossible for someone to believe inconsistencies. Both knowledge and belief are closed under logical inference. All valid formulae are 'known'. Aumann's semantics [Aum] uses partitions rather than Kripke structures but is known to be equivalent [FHMV] and suffers from the same difficulty. Various researchers including ourselves have looked into this issue [Pa87, P94a, FHMV, Stal99, Gaif04]. There is also more recent work by Artemov [Art04], Artemov and Nogina [AN05], and by Fitting [Fit04], aimed towards the Logic of Proofs.

The proposal we make here draws on this previous work just cited, as well as on the work of Ramsey [Ram26], Hayek [Hay36], and the currently influential philosophical streams of functionalism and contextualism.

Consider Ramsey. Imagine that Carol assigns probabilities of .2, .4, and .8 respectively to events

$X, Y, X \cup Y$. One could say that these probabilities are *inconsistent*. But in fact nothing prevents Carol from accepting bets based on these probabilities. What makes them inconsistent from Ramsey's point of view is that we can make Dutch book against Carol – i.e., place bets in such a way that no matter what happens, she will end up losing money. Thus inconsistent beliefs, on Ramsey's account, are *possible*, but unwise.

Hayek [Hay36] considers an isolated person acting over a period according to a preconceived plan. The plan "may, of course, be based on wrong assumptions concerning external facts and on this account may have to be changed. But there will always be a conceivable set of external events which would make it possible to execute the plan as originally conceived". So one could relate a person's beliefs, not to some Kripke structure but to a plan – and assume that the beliefs must be such as to make the plan workable from the agent's point of view. If I take my umbrella, it does not follow that it *is* raining, but it does follow that I believe it is raining (or will rain soon). No Kripke structure is needed here at all, just some understanding of who I am and what my plans are.

Functionalism [Block80] is in fact a philosophy of mind which relates the mind to behavior.[1] Block says, "The functionalist answer to 'What are mental states?' is simply that mental states are functional states." This stance can clearly be reconciled with Ramsey and Hayek, because instead of thinking of an agent's beliefs and knowledge as related to some Kripke structure and *causing* his behaviour, we can think of the behaviour itself as a clue to (or, more bravely, even as defining) the knowledge or belief.

As for *contextualism*, its original purpose is to address skeptical doubts. Moore said that I know that I have hands and that therefore I know that there is an external world. However, the existence of the external world is not proved quite so simply. The contextualist answer is to say that when I say that I have hands, I am operating in a different context from the one where I doubt that the external world exists. In conventional Kripke structure semantics, the context appears as the Kripke structure itself. We evaluate an agent's knowledge *within* a Kripke structure; we do not ask *how* the agent knows that that is the right Kripke structure.

We start with a group of questions raised earlier and then propose a formalism which addresses these questions in a way which does justice to our pre-theoretic intuitions.

1. Ravi knows all the axioms of Peano Arithmetic. Does Ravi know all the theorems?

2. Ravi knows all the axioms of Peano Arithmetic. He also knows that all true existential formulas are provable in PA. If some universal formula (say Goldbach's conjecture GB) is true, then its negation is not provable in PA. Does Ravi know that $\neg GB$ is not provable in PA and hence that GB must be true?

[1]The important difference between functionalism and behaviorism will not be too relevant to us here.

3. A dog is digging in your garden. There is indeed a bone buried where he is digging. Does the dog know that there is a bone buried in your garden?

4. Do fish know that they are in water?

5. Lois Lane knows that Superman flies. Clark Kent *is* Superman. Does Lois Lane know that Clark Kent flies?

6. Jill knows that Germany is north-east of France. Jill knows that Berlin is in Germany. And Jill also knows simple geometric facts about north, south, etc. Does Jill necessarily know that Berlin is north-east of France?

In this paper we shall introduce two notions of knowledge, e-knowledge (which is propositional) and i-knowledge (which is sentence-based and only appropriate for language users). These two notions will be used to sort out various knowledge puzzles.

2 The Setting

In our setting we imagine an observer o who is pondering on what some agent i knows. We assume (for convenience) that o thinks of a proposition expressed by a sentence as a set of possible worlds where that sentence is true, but that the observee i need not even have a language or a notion of truth. However, it is assumed that i does have some plans. Even if i is just a dog digging for a bone, o understands that i has a plan and roughly what that plan is. And we shall use this plan to make it possible for o to attribute beliefs and knowledge[2] to i.

We also assume that there is a context C which is the set of *relevant* possible worlds, and that worlds outside C, even though they are there, are not considered in deliberating about i's knowledge or beliefs. It is not assumed that i knows that there are worlds outside C; in some sense i lives inside C, but we will assume that o does know that there are possible worlds outside C. The purpose of the context is to avoid considering cases which are possible but strange, like the laws of physics failing, or Bush suddenly joining the Green party. A plan is assumed to work in *normal* circumstances, and an agent i is only required to have the sort of knowledge which would be enough for the plan to be carried out in normal circumstances.[3]

So let P be i's plan at the moment, and let $\pi(P)$ be the set of worlds w in C such that the plan is *possible* at w.

[2] Of course o might on occasion be wrong about the plans. For instance, if i is a bird, then o may believe that i sits on an egg because she is overcome by maternal feelings. But research has shown that actually the bird is trying to cool its behind which gets hot at just the right time, and the egg is a nice, heat-absorbing object. But even in such a case, i does have a plan, namely to cool its behind, and we can apply our theory of knowledge to i in view of this information.

[3] Note that with conventional Kripke semantics, when we calculate the agents' knowledge in a Kripke structure, the Kripke structure is such a context. It is assumed to be 'known' but not in any explicit way.

There are two senses in which the plan may be possible. One is that the plan can actually be carried out. For instance for the dog digging for a bone, that possibility means that the ground is soft enough. The other sense is that the plan actually yields i the benefit which i wants; in this case, the actual presence of a bone. So we can think of a plan P either as just a bare-bones algorithm (sorry!), or as an algorithm supplemented by a utility function.[4] The theory can be applied with either notion of a plan, but the one which includes the utilities is more informative.

So formally,
$\pi(P) = \{w|\ w \in C \wedge w\ enables\, P\}$.

Let ϕ be a sentence. Then $||\phi|| = \{w|w \models \phi\}$, the set of worlds[5] where ϕ is true, is the *proposition* corresponding to the sentence ϕ. Note that if ϕ and ψ are logically equivalent, then $||\phi|| = ||\psi||$. Both sentences and propositions will enter into our semantics.

Definition: We will say that i *e-believes* ϕ, $B^i_e(\phi)$ if $\pi(P) \subseteq ||\phi||$. [6] And we will say that i *e-knows* ϕ, $K^i_e(\phi)$ if i e-believes ϕ, the real world w satisfies ϕ, and in all similar cases, the agent would e-believe ϕ only if ϕ were true. We will suppress the superscript i when it is clear from context.

The choice of 'similar' cases is left to o, so that if o were to decide that only *this* case is similar, then mere truth plus e-belief would be enough to justify e-knowledge.[7]

Note that we are applying these two operators only to objective propositions and are not looking at iterated applications of the two operators. But within this limitation, both operators B_e, K_e are normal and K_e is veridical.[8]

It is obvious in terms of the semantics which we just gave that the statements "The dog e-believes that there is a bone where he is digging", and in case there *is* a bone, "The dog e-knows that there

[4]For simplicity, we will assume that this function u is real-valued etc, but it will usually not matter. After all, you *can* manage to set up an experiment which will offer a dog a choice between a steak and a walk, but you can hardly offer it a lottery which is .7 steak and .3 walk. Thus the dog's utility function is at best an ordering and lacks the additive structure which the standard theory assumes.

[5]For simplicity we shall assume that in case i is actually a person, then o, i have the same notion of possible world. Puzzles like Kripke's Pierre puzzle will be addressed in another place. One way to deal with such situations is to replace *possible worlds* by *MCT's; maximal consistent first order theories.* There is such an MCT in which Londres is beautiful and London is not, even though there is no possible world in which this can happen. Thus suppose Pierre is planning to buy a train ticket for a trip from Londres to London, then there is no possible world in which he can do this. But there are MCT's where he can, and if o is aware of Pierre's deficiency, he can still understand what Pierre believes. See [Ye99].

[6]Mike Levin has suggested the locution "*i pre-supposes* $||\phi||$".

[7]We are deliberately simplifying an account which, for a more realistic treatment, needs to be more sophisticated. For instance, if we think of a plan P as (deterministic) Dynamic Logic program, and ψ is the agent i's desired end condition, then we can say that the agent believes $< P > \psi$. One could go further and say that if the agent has a choice of several plans $P_1, ..., P_k$ and uses P_1, then we can attribute to the agent the belief that the expected utility of P_1 is largest among the various P_i. But such details are not appropriate for a first incursion into this subject.

[8]We have not gone into the issue of justification of knowledge for the following reason: we know from Gettier that justifications are slippery and there is no such thing as the ultimate justification which simply cannot be undermined. Given this we might as well allow the weakest justification for knowledge, namely truth. We will consider stronger notions of justification in another place.

is a bone where he is digging", are both going to be true. If the dog is digging, but o knows there is no bone, then o is also justified in saying, "The dog falsely e-believes that there is a bone where he is digging".[9]

We will also have the laws, $K_e(\phi) \to \phi$, and rules like

$$\frac{\phi \to \psi}{K_e(\phi) \to K_e(\psi)}$$

Thus a lot of logical omniscience goes along with e-knowledge, but only within the context of a single plan. For instance, a white employer in the process of interviewing candidates may hold the e-belief, "All races are equal", but hold a different e-belief in the context of someone wanting to date his daughter. If he is aware of the two incompatible stances, then he is a hypocrite, but it is quite possible that he is not so aware. *Or*, one may drive a colleague to the airport for a two week vacation in Europe and then forget, and arrange a meeting three days later at which this colleague's presence is essential. But within the context of a single (short) plan, consistency and logical omniscience will tend to hold. The situation is more complex with multiple plans. And there is nothing to prevent an agent from having one e-belief in one plan and another contradicting e-belief in another plan. But pragmatic considerations will encourage the agent i to be consistent and to use logical closure.

Suppose someone has a plan P consisting of, "If ϕ then do α, else do β" and another plan P' consisting of "If ϕ then do γ, else do δ". Now we find him doing α and also doing δ (we are assuming that the truth value of ϕ has not changed). We could accuse him of being illogical, but there is no need to appeal to logic. For he is doing Dutch book against himself.

Presumably he assumed that $u(\alpha|\phi) > u(\beta|\phi)$ but $u(\alpha|\neg\phi) < u(\beta|\neg\phi)$. Thus given ϕ, α was better than β but with $\neg\phi$ it was the other way around. Similarly, $u(\gamma|\phi) > u(\delta|\phi)$, but $u(\gamma|\neg\phi) < u(\delta|\neg\phi)$. And that is why he had these plans. But then his choice of α, δ results in a loss of utility whether ϕ is true or not. If ϕ is true then he lost out doing δ and if ϕ is false, then he lost out doing α.

For a concrete example of this, suppose that on going out I advise you to take your umbrella, but fail to take mine. If it is raining, there will be a loss of utility for I will get wet. If it is not raining, there will be a loss of utility because you will be annoyed at having to carry an umbrella for no good reason. My choice that I advise you to take your umbrella, but fail to take mine, is not *logically* impossible. It just makes no pragmatic sense. A similar argument will apply if someone 'believes' ϕ, 'believes' $\phi \to \psi$ and disbelieves ψ. If such a person makes plans based on assuming these three conditions, then he will make choices which do not maximise his utility. Of course such arguments

[9] It is possible to create Gettier cases even for dogs. Suppose for instance that Fido buried a bone in spot S, and later on o goes and digs out that bone and replaces it with another one. The dog comes back and starts digging at spot S. We could now say, "The dog believes there is a bone, and in fact there is a bone, but the dog does not know there is a bone". But we shall not go into this issue as we are not discussing justification here.

go back to Ramsey [Ram26] and Savage [Sav54].

If we assume that over time, people learn to maximise their utility (they do not always but often do), then they will 'learn' a certain amount of logic and they will make certain obvious logical inferences.

A little girl who was in kindergarten was in the habit of playing with her older sister in the street every day when she came home. Once, her older sister was sick. So the little girl went to visit her sick sibling in her bedroom, and then, as usual, went out into the street to play with her. Clearly the little girl had not yet learned to maximise her utility!

But a lack of "logic" may not necessarily do harm. There once was a beggar who would always take the $5 bill if offered a choice between $5 and $10. Many people tried this, and he consistently took the $5 on every occasion. Finally when someone asked him why he did this since $10 was clearly more, he replied, "If I even once took the $10, no one would give me the $5 any more." Here the action of taking the $5 bill makes sense since it is a repeated game and the actions of donors are influenced by the beggar's own choices.

3 A second notion of knowledge

We now define a second notion of knowledge which does *not* imply logical omniscience. This is a more self-conscious, language-dependent notion of knowledge.

For agents i who do have a language (assumed to be English from now on), their plan may contain linguistic elements. At any moment of time they have a finite stock of currently believed sentences. This stock may be revised as time passes. These agents may perform atomic actions from time to time, and also make observations which may result in a revision in their stock of believed sentences.

Thus Lois seeing Superman in front of her will add the sentence "Superman is in front of me", to her stock, but, since she does not know that Clark Kent is Superman, she will *not* add the sentence "Clark Kent is in front of me". Someone else may add the sentence "I see the Evening Star", but not the sentence "I see the Morning Star" at 8 PM on a summer night. A person who knows that $ES = MS$, may add the sentence, "Venus is particularly bright tonight." In any case, this stock consists of sentences and not of propositions.

The basic objects in the agents' *plans* are atomic actions and observations which may be active (one looks for something) or passive (one happens to see something). These are supplemented by the operations of concatenation (sequencing), *if then else*, and *while do*, where the tests in the *if then else* and *while do* are on *sentences*. There may also be recursive calls to the procedure: *find out if the sentence ϕ or its negation is derivable within the limits of my current resources, from* *my current stock of beliefs.* Thus if i's plan has currently a test on ϕ, then, to be sure, the stock of sentences will be consulted to see if ϕ or its negation is in the stock. But there may also be a recursive call to a procedure for deciding ϕ. If someone asks "Do you know the time?", we do not usually say, "I don't", but look at our watches. Thus consulting our stock of sentences is typically only the first step in deciding if some sentence or its negation can be derived with the resources we have.

This difference between sentences and propositions matters as we now show. Suppose for instance that Lois Lane has invited Clark Kent to dinner but he has not said yes or no. So she forms the plan, *While I do not have a definite answer one way or another,* if *I see Clark Kent, I will ask him if he is coming to dinner.* Here *seeing Clark Kent* is understood to consist of an observation followed by the addition of the sentence "I am seeing Clark Kent" to her stock.

Suppose now that she sees Superman standing on her balcony. She will *not* ask him if he is coming to dinner as the sentence "I am seeing Clark Kent" will not be in her stock of sentences. And this is the sense in which she does *not* know that when she is seeing Superman, she is also seeing Clark Kent. If she *suspects* that Clark Kent is Superman, then it may happen that her recursive call to the procedure "decide if I am seeing Clark Kent" will take the form of the question, "Are you by any chance Clark Kent, and if so, are you coming to dinner?" addressed to Superman.

Have we given up too much by using sentences and not propositions as the objects of i-knowledge? Suppose her plan is, "If I see Jack and Jill, I will ask them to dinner" but she sees Jill first and then Jack so that she adds the sentence "Jill and Jack are in front of me" to her stock. Will this create a problem? Not so, because if a sentence ϕ is in her stock, ϕ easily implies ψ, and she needs to know the value of ψ so she can choose, then the program *find out about* ψ which she calls will probably find the sentence she needs. If the program terminates without yielding an answer she may well have a default action which she deems safest.

So here we make use of the fact that Lois does have a reasonable amount of intelligence. Even if she does not explicitly add some sentence ψ to her stock, when she comes to a point where the choice of action depends on ψ, she *will* ask if ψ or its negation is derivable from her present stock of sentences, possibly supplemented by some actions which add to this stock.

Definition: If an agent a comes to a point in her plan where her appropriate action is *If* ϕ *then do* α *else do* β, and she does α, then we will say that she *i-believes* ϕ. If, moreover, ϕ is true, and we believe that in a similar context she would judge it to be true only if it *is* true, then (within the context of this plan) we will say that she *i-knows* ϕ.

A common example of such a plan is the plan to answer a question correctly. Thus if an agent is asked "Is ϕ true?", the agent will typically call the procedure "decide if ϕ is true", and then answer "yes", "no", or "I don't know" in the appropriate cases.

Now note that if an agent means to deceive, then the same procedure will be called, but the answers given will be the opposite of the ones indicated by the procedure. But if we ourselves know that the agent's plan is to deceive, then we can clearly take the statement "ϕ is false" to indicate that the agent believes ϕ.

We no longer have the law that if the agent i-knows ϕ and ϕ implies ψ then the agent of necessity i-knows ψ. But if the agent has the resources to decide ϕ and the proof of ψ from ϕ is easy, then she might well also know ψ. But her notion of "easy" may be different from ours, and how much effort she devotes to this task will depend on her mood, how much energy she has, etc.[10]

Note now that we often tend to assign knowledge and beliefs to agents even when they are not in the midst of carrying out a plan. Even when we are brushing our teeth we are regarded as knowing that the earth goes around the sun. This can be explained by a continuity assumption. Normally when we *are asked* if the earth goes around the sun or vice versa, we say that the former is the case. An agent is then justified in assuming that even in between different occasions of being so asked, if we *were* asked, we would give the same answer. This assumption, which is usually valid, is what accounts for such attributions of knowledge.[11]

It has been suggested in this context (e.g. by Stalnaker [Stal99]) that such issues can be addressed by using the notion of fine grain. By this account, if I understand this correctly, logical equivalence is too coarse a grain and that a finer grain may be needed. So if two sentences are in the same fine grain and an agent knows or believes one, then the agent will also know or believe the other. But if we try to flesh out this metaphor then we can see that it is not going to work.

For instance if $G(\phi, \psi)$ means that ϕ and ψ are in the same grain, then G will be an equivalence relation. But surely we cannot have transitivity in reality, because we could have a sequence ϕ_1,...,ϕ_n of sentences, any two successive ones of which are easily seen to be equivalent, whereas it is quite hard to see the equivalence of ϕ_1 and ϕ_n.

Moreover, "being in the same grain" sounds interpersonal. If two molecules are in the same rice grain for you, then they are also in the same fine grain for me – it is just a fact of the matter. But in reality people differ greatly in their ability to perceive logical equivalences.

Thus suppose some set theorist thinks of some new axiom ϕ and wonders if ϕ implies the continuum hypothesis, call it ψ. The set theorist may find it quite easy to decide this question even if we see

[10]For instance if a customer sits down on a bar stool and the bartender sees him, we do not need to ask, "Does the bartender know he wants a drink?" Of course he knows. But suppose a woman suffering from a persistent cough calls her husband and says, "I got an appointment with the doctor for 2:30 PM", she may later think, "I wonder if he realized that I cannot pick up our daughter at 3". When i knows ϕ and ψ is deducible from ϕ, then whether we assume that i knows ψ will depend less on some objective distance between ϕ and ψ than on what we know or can assume about i's habits.

[11]There is a rather poignant story about Ronald Reagan in his last days. On a visit to his doctor he saw a reproduction of the White House, picked it up, and would not put it down. When being asked to explain his behavior he said, "I know it has something to do with me, but I don't know what."

no resemblance between ϕ and ψ. And if he does not find it easy, he may look in the literature or ask another set theorist, processes which cannot easily be built into a formal theory. And they should not be! For if they were easy to build into a formal theory, then the representation is almost certain to be wrong.

Or a chess champion may be able to see 20 moves (40 half-moves) ahead in the end game, but you and I cannot. And he too usually cannot do this during the middle game. Thus context, habit and expertise matter a lot.

Many of us have learned the theorem that an isosceles triangle must have equal corresponding angles. If AB = AC in the triangle ABC then the angles at B and C must be the same. The proof I learned used a perpendicular line drawn from the point A to the line BC.

But a shorter proof was discovered recently, I believe by a computer. The triangles, ABC and ACB are *congruent!*. Ergo, the angles B and C must be the same!

That this proof was not found earlier has nothing to do with any realistic notion of distance. we just *did not know it!* Knowledge is often the result of a creative insight, and it is a pity that previous models of knowledge have left out this crucial feature of knowledge.

Going back to our original question, we do now have the fact that if Ravi i-knows the axioms of Peano Arithmetic, he need not i-know the theorems. But he does e-know them, as they are true in all possible worlds. The presence of these two notions, each of which captures part of our intuition makes it possible to get out of certain quandaries. As for the fish, since "in the water" is their context, they do know they are in water, but this proposition is indistinguishable for them from any other universally true proposition.

4 Common Knowledge:

In [CM81], Clark and Marshall point out that common knowledge seems to be needed even for such a simple event as Ann saying to Bob, "Have you ever seen the movie showing at the Roxy tonight?" In this case the movie showing at the Roxy is *Monkey Business*, but an earlier (incorrect) announcement had said that the movie was *A Day at the Races*. Clark and Marshall show that for Bob to understand Ann properly it must be *common knowledge* (or mutual knowledge) between the two of them that the movie shown is indeed *Monkey Business (MB)*, and that a simple condition like both knowing this fact is not enough for the reference to go through.

For instance, suppose that they read the morning newspaper together and it said that *A Day at the Races* would be playing. Later they *separately* read the correction that it would be *Monkey Business* which would be playing. When Ann asks Bob, "Have you ever seen the movie showing at the Roxy tonight?", Bob is likely to think that she is asking about *A Day at the Races*. Thus it is necessary for him to know that she knows about *Monkey Business* and that she knows that he knows. But even these two assumptions do not actually suffice, as more complex examples show – mutual knowledge is needed.

But mutual knowledge, that *both know that movie* = MB, *both know that both know that movie* = MB, etc, is an infinite and therefore implausible conjunction. This is a problem, because mutual knowledge does seem necessary. Clark and Marshall conclude, "Mutual knowledge is an issue we cannot avoid. It is likely to complicate matters for some time to come."

We proceed to address this problem in a way analogous to the one we used earlier. For a single agent to undertake a plan, the agent must believe that the conditions are such that the plan can be carried out. Or if we are talking e-belief, then the fact that the agent is proceeding with this plan can be taken by us (o) to be a criterion for the claim that the agent believes such conditions exist. Instead of seeing the belief as the agent's reason for undertaking the plan, we reverse gears and take the attempted execution of the plan to be our criterion for the belief that the world is such that the plan is executable.

We can now use this method for shared action. Suppose than Ann and Bob are dancing and at the moment when Bob puts his right foot forward, Ann must pull her left foot backward. If one action happens without the other then either Bob will fall, or he will step on Ann's foot, both of which are undesirable occurrences.

But does Bob ask himself, "Does Ann know that I know that she knows that I know that she is going to pull her left foot backward?" Of course not. What does happen is that they are both listening to the same music and if they have already trained themselves to dance properly to this music, then their movements will be in concord, and co-ordinated action will take place without *conscious* common knowledge (and is there any other kind?). If they are not trained to dance the particular dance, then no amount of common knowledge will help them.

To understand what happens, we resort to a hypothesis of *methodological solipsism.* If Bob were dancing with a mechanical woman who had been programmed in a certain way to move in accord with the music, then Bob could dance with the mechanical Ann *in the same way* as he would dance with a real Ann. In other words, all that actually concerns Bob are Ann's actual movements, and not how these movements arise from her beliefs and utilities.

Clearly, if Bob is dancing with a mechanical woman, then no common knowledge is needed for successful dance – all that Bob needs to do is to dance *to the music.* If Bob *is* dancing with a mechanical Ann but is not aware of it, then what he thinks is common knowledge that they are dancing the cha-cha will in fact amount to a *successful* delusion on his part.

Thus for a dance, or for another co-ordinated action which appears to require common knowledge, all that is actually needed for this action to succeed is that the individual parts of the action be co-ordinated, and this can be brought about via means other than common knowledge.

Common knowledge appears not to be *necessary.* Is it at least sufficient? There again the answer is no.

For imagine a group of prisoners who are guarded by two guards at the prison gate. They dare not rush either of the guards because the other guard will then shoot the first prisoner who goes forward.

Then one day the prisoners realize that one guard leaves for lunch at noon, and his replacement arrives only at 12:15. Two of the prisoners then decide to simultaneously rush the lone guard at 12:05, they pin him down, and then they all escape.

But why did they need to make the plan at all? Surely the common knowledge that the guard could be rushed at 12:05 existed all along! But the plan was needed *in addition to* the "common knowledge."

In ordinary life, such ready-made joint actions already exist in our 'library' of possible joint actions. Dating is a joint action which can now take place when it is a custom, but before it was a custom, it did not take place, even though the conditions for the knowledge *dating is possible* existed all along. Was the knowledge there? My intuition is that it did not for the most part – it simply did not occur to people. Other social procedures, lectures, elections, committee meetings, have all been *invented.* Once they exist, we can take advantage of our knowledge of their existence, and that others know about them, and then carry them out. Thus arranging a lecture may involve finding a speaker, arranging a room, and then creating something like common knowledge of the lecture in the target audience. But in doing all this we make use of a pre-existing procedure.

Does that mean that conventional (Kripke) common knowledge has no role to play? Well, conditions which 'enable common knowledge' according to the standard theory will be needed for procedures to work, even when common knowledge is not itself present. Suppose for instance that when Ann and Bob are dancing, certain beats can be heard only by Ann and not by Bob, whose ears hear fewer frequencies. However, they are not aware of this. In that case it is quite possible that one of them will fall because on a beat which only Ann hears, she will move and Bob will not.

We *could* say that common knowledge failed, but in fact what failed was Bob's hearing.

In [Pa91] we pointed out the following 'method' solution to the muddy children puzzle. Suppose the children are told, "Wait until all the muddy children you can see have already said 'I don't know'. If your turn comes then, and no one has yet said 'My foreheard is muddy', then say about yourself 'My forehead is muddy' ". It is easy to see that this procedure gives exactly the same answer as the standard one based on the Kripke model notion of knowledge, and is sound. Moreover, it can be learned even by five year old children who cannot possibly master the logic of knowledge.

Thus the solution to the common knowledge problem is that *society* sets up certain procedures which work when the physical circumstances which *would* justify common knowledge are present. Then, we are *trained* to act in a certain way when we know we are taking part in such a procedure. These different trainings, together with the physical circumstances needed for common knowledge, create co-ordinated action *without common knowledge being actually present.*

5 Other levels of knowledge:

It is well known[12] [Pa86, PK] that there are many levels of knowledge short of common knowledge. Do these levels also 'arise' in social situations? In other words, are there social situations where the physical setup would justify something short of common knowledge, and that is sufficient for co-ordination? And how do such situations arise?

In [CM86], Chandy and Misra consider knowledge gained (and lost) thorugh asynchronous messages and prove the following result. Suppose that h and h' are sequences of events, including messages sent and received and that h is an initial segment of h'. Moreover, after h, process A does not know fact P, whereas after h', C knows that B knows that A knows P. Then there must have been a sequence of messages starting with A, passing through B (possibly via X,Y,Z), and then to C.

Thus it looks as if $K_C K_B K_A(P)$ can only be created over time. A formula like $K_E K_D K_C K_B K_A(P)$ with more K operators should take even more (at least four) messages. But in fact this apparent requirement does not hold generally, and the assumption of asynchronicity in Chandy and Misra's theorem was crucial. Otherwise, such a complex state of knowledge can be created in a single step!

Consider the following protocol. Nature is capable of sending simultaneous signals to A, B, C respectively. One possible signal s is $< a, b, c >$ meaning that A receives a, B receives b and C receives c. The other possible signals are $< a, b, x >, < a, y, x >$, and $< z, y, x >$. All the signals containing a are sent only if P is true. Initially P is false, but it becomes true and nature sends the signal $< a, b, c >$. It is easy to see that this single signal creates the situation $K_C K_B K_A(P)$ whereas before the signal, A did not know P (it was in fact false). Moreover, we do not overshoot our target. Common knowledge of P is not created as A does not know that B received a b and B does not know that C received a c. So A does not know that B knows about P and B does not know that C knows about P, let alone about $K_B K_A(P)$.

Theorem: Let Σ be a state of knowledge among a finite number of agents $1, ..., n$ involving propositions $P_1, ..., P_m$. Assume that the agents initially know nothing about the P_i or each other's knowledge of the P_i. Assume moreover that Σ can be represented by a finite Kripke structure. Then the state of knowledge Σ can be produced by two announcements (by Nature, God or father), the

[12] When we use the word "knowledge" in this section we shall be referring to the standard Kripke structure account of knowledge. We shall use the appelation "k-knowledge" when we need to be quite explicit about this.

first public and the second individualized.

Proof: For notational simplicity, assume that there are only two agents. Moreover, as we know from Aumann, the finite Kripke structure can also be represented as a pair of partitions $\{X_1, ..., X_k\}$ and $\{Y_1, ..., Y_p\}$ on some finite space W plus a valuation V mapping $W \times \{P_1, ..., P_m\}$ into $\{0,1\}$. Then Nature first announces publicly what V is and what the X_i, Y_j are; that it will then choose some world $w \in W$, and, in the second message, it simultaneously tells the first agent the X_i in which w falls and tells the second agent the Y_j in which w falls. These two announcements will then achieve the required state of knowledge Σ. □

The muddy children puzzle is an example of such a situation as each child receives a different signal, each signal being compatible with two states of nature, one where it is itself muddy and one where it is not. But if there are n children, then there are $n \cdot 2^{n-1}$ possible signals, and the father's announcement, "At least one of you is muddy", reduces the number by one. In conventional Kripke terms, there is now common knowledge that some child is muddy and such common knowledge was not there before. But as we saw, a different, procedural account of what happens is perfectly possible.

Do situations involving complex states of knowledge which are *not* common knowledge arise? And are they used in procedures? I believe so. The game of bridge, where each player receives an individual signal (his own cards plus the public bidding) is such an example. The TV series *Yes, Prime Minister* is a wonderful example of such complex knowledge situations in real life.[13]

But our main point was that knowledge does not necessarily *exist* merely because the conditions for it to exist are present. Rather, we have socially available co-ordinated procedures which we have learned through being members of society, and the physical conditions which *would justify* Kripke structure knowledge are needed for these procedures to work. But the knowledge itself, as a psychological state, is neither necessary nor sufficient.

6 k-Knowledge and i-knowledge

If we refer to the conventional account of knowledge as k-knowledge, then we should notice that it is in fact a special case of i-knowledge. For instance suppose Ravi has been told the axioms of Group theory and in the process of taking an examination, he wants to decide if some formula ϕ is true in the context of these axioms. He does have two ways of implementing the procedure *decide if ϕ is true.* He can either run through all possible groups in his mind and see if ϕ is true in all of them. Alternatively, if ϕ is first order, he can run a proof procedure to derive it from the axioms.

[13] In one incident, Sir Humphrey insists that the Trident missile project be retained. When asked why, he explains that the project cannot actually defend Britain, but that it *will create* the belief that Britain is defended. When asked if it is the Russians who need this belief, he says no, for the Russians *know* that Britain is not defended by the project – it is the British public which needs this illusion.

Thus both of these implementations of *decide if ϕ is true* can be seen as special cases relevant to i-knowledge. However, i-knowledge is a more general notion and captures our intuitions better. We should also suggest that it is more realistic than a complexity-theoretic notion. For the latter still conflates, *It is possible to find out if ϕ in a reasonable amount of time* with *It has been found out that ϕ.* i-knowledge chooses the latter and rightly so. For if some mathematician proves some difficult theorem with a lot of effort, surely we do not say that he does not know it because the effort was great. Sometimes we know something only after a great deal of effort, and then we do know it. Sometimes we could find out something with only a small amount of effort but we fail through laziness or bad luck. And then we do not know it, even if we *should have.* i-knowledge makes the right choice in both cases.

In the one-person case we would normally expect that if the Kripke structure in question is clear enough, then i-knowledge will be a subset of k-knowledge. In other words, nothing can be i-known unless it is k-known. But this need no longer hold with more than person. For instance if the Continuum Hypothesis is true, then it is common knowledge. And so if i-knowledge were a subset of k-knowledge, it would impossible for A to know that B does not know it. With i-knowledge, if A has some idea of the algorithm which B is using to decide ϕ and knows already that no such algorithm will work, then A can very well know that B does not know ϕ, even though in terms of k-knowledge, ϕ, if true, might be common knowledge.

Acknowledgements: We thank Sergei Artemov, Samir Chopra, Horacio Arlo Costa, Haim Gaifman, Mike Levin, Larry Moss, and Eric Pacuit for comments. Thanks to Karen Kletter for bringing the relevance of *Yes, Prime Minister* to our attention. The information about chess came from Danny Kopec. This research was supported by a grant from the PSC-CUNY faculty research assistance program.

References

[AN05] Artemov, S., and E. Nogina, "Basic epistemic logics with justifications", *Technical report TR-2005004*, CUNY Ph.D. program in Computer Science, 2005. To appear in TARK-X.

[Art04] Artemov, S., "Evidence based common knowledge", *Technical report TR-2004018*, CUNY Ph.D. program in Computer Science, to appear in *TCS*.

[Aum] Aumann, R., "Agreeing to disagree", *Annals of Statistics*, **4** (1976) 1236-1239.

[Block80] Block, N., "Introduction: what is functionalism?", in *Readings in Philosophy of Psychology*, (Ned Block, editor) Harvard University Press (1980) 171-184.

[CM81] Clark, H., and C. Marshall, "Definite reference and mutual knowledge", in *Elements of Discourse Understanding*, Ed. Joshi, Webber and Sag, Cambridge U. Press 1981, 10-63.

[CM86] Chandy, M., and J. Misra, "How processes learn", *Distributed Computing*, **1**, (1986) 40-52.

[CW05] Chopra, S., and L. White, "Attribution of Knowledge to Artificial Agents and their Principals", *Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence*, Edinburgh, August 2005, forthcoming.

[D] F. Dretske, *Knowledge and the flow of information*, MIT Press, 1981.

[FHMV] Fagin, R., Halpern, J., Moses, Y. and Vardi, M., *Reasoning about knowledge*, M.I.T. Press, 1995.

[Fit04] Fitting, M., "A logic for explicit knowledge", to appear in proceedings of *Lgica 2004*.

[Gaif04] Gaifman, H., "Reasoning with limited resources and assigning probabilities to arithmetical statements", *Synthese*, **140** (2004) 97-119.

[Hay36] Hayek, F.A., *Individualism and Economic Order*, University of Chicago Press. See especially chapters II and IV.

[HF] J. Halpern and R. Fagin, "A formal model of knowledge, action and communication in a distributed system", in *ACM PODC 1985*, pp 224-236.

[HM] J. Halpern and Y. Moses, "Knowledge and common knowledge in a distributed environment", in *ACM PODC 1984*, pp 50-61.

[KPN] Krasucki, P., Parikh, R., and Ndjatou, N., "Probabilistic Knowledge and Probabilistic Common Knowledge" *ISMIS 90*, North Holland 1990, 1-8.

[Noz81] Nozick, R., *Philosophical Explanations*, Harvard University Press, 1981.

[Pa86] Parikh, R., "Levels of Knowledge in Distributed Computing", in *Proc IEEE Symposium on Logic in Computer Science,* June 1986, 314-321.

[Pa87] Parikh, R., "Knowledge and the Problem of Logical Omniscience" *ISMIS- 87* (International Symp. on Methodology for IntelligentSystems), North Holland (1987) 432-439.

[Pa91] Parikh, R., "Finite and Infinite Dialogues", in the *Proceedings of a Workshop on Logic from Computer Science*, Ed. Moschovakis, MSRI publications, Springer 1991 481-498.

[Pa91b] Parikh, R., "Propositions, propositional attitudes and belief revision" in K. Segerberg, M. Zakharyaschev, M. de Rijke, H. Wansing, editors, *Advances in Modal Logic, Volume 2*, CSLI Publications, 2001.

[P94a] Parikh, R., "Logical omniscience" in *Logic and Computational Complexity*, LNCS 960, 22-29.

[P95] Parikh, R., Knowledge based computation (Extended abstract), in *Proceedings of AMAST-95*, Montreal, July 1995, Edited by Alagar and Nivat, Lecture Notes ib Computer Science no. 936, 127-42.

[Pa95b] Parikh, R., “Logical omniscience”, in *Logic and Computational Complexity* Ed. Leivant, Springer Lecture Notes in Computer Science no. 960, (1995) 22-29.

[Pa02] Parikh, R., “Social Software”, *Synthese*, **132**, Sep 2002, 187-211.

[P03] Parikh, R., Levels of knowledge, games, and group action, in *Research in Economics*, **57**, (2003) 267-281.

[PK] Parikh, R., and P. Krasucki, “Levels of knowledge in distributed computing”, *Sadhana - Proc. Ind. Acad. Sci.* **17** (1992) 167-191.

[PaPa] Pacuit, E., and R. Parikh, A Logic for communication graphs, to appear in the proceedings of *DALT 2004*.

[PR'85] Parikh, R., and R. Ramanujam, Distributed processes and the logic of knowledge, in *Logic of Programs*, LNCS #193, Springer 1985, 256-268.

[PR'03] Parikh, R., and R. Ramanujam, A Knowledge based Semantics of Messages, in *J. Logic, Language and Information*, **12**, (2003) 453-467.

[Ram26] Ramsey, F.P., ‘Truth and probability’, in *The Foundations of Mathematics*, Routledge and Kegan Paul (1931), 156-198.

[Sav54] Savage, Leonard, *The Foundations of Statistics*, Wiley 1954.

[Stal99] Stalnaker, R., *Context and Content*, Oxford University Press, 1999.

[Witt] Wittgenstein, L., *Philosophical Investigations*, MacMillan, 1958.

[Ye99] Ye, Ruili, *Belief, Names and Modes of Presentation*, Doctoral dissertation, CUNY 1999.