Logical Omniscience in the Many Agent Case

Rohit Parikh

City University of New York

July 25, 2007

Abstract: *The problem of logical omniscience arises at two levels. One is the individual level, where an agent is assumed to have reasoning powers which are unrealistic. The other, equally important one, is where two or more agents are supposed to share a state of knowledge (perhaps* common knowledge*) which is read off from a physical situation, but which may not hold in practice.*

By reducing knowledge to strategies in games (rather than the other way around) we show how to get around this problem.

Essentially, we are using the same trick which Ramsey used when he derived subjective probability from an agent's choices rather than the other way around.

Preamble: Consider the following scenario. Ann is sitting on a chair in front of which there is a vase with a dozen roses in it. Bob can see both Ann and the roses. Charlie can see Ann and Bob and the roses.

We could now ask: *Does Ann know p?* where p = There are roses in front of her. I.e., $K_a(p)$? Does Bob know that she knows? ($K_bK_a(p)$?) Does Charlie know that Bob knows that Ann knows ($K_cK_bK_a(p)$) ?

Both common sense and the corresponding Kripke structure tell us that the answer to all three questions is *yes*. Indeed if they can see each other then p is *common knowledge* among them.

Let us now change the meaning of p. In this new example, Ann, Bob and Charlie are all as before, but what is in front of Ann is not a vase of roses, but a blackboard with the number 1243 written on it. Let p now denote the fact that the number on the blackboard is composite.

Logically the situation is not changed. Since 1243 is composite (113 times 11), this is a necessary truth, Ann knows it, Bob knows that Ann knows it, and Charlie knows that Bob knows that Ann knows it.

But are *we* sure that this is the case? It could be that Ann finds numbers greater than 100 to be a mystery. Or perhaps she *is* actually a number theorist but sexist Bob thinks that she is number-challenged. Or perhaps Bob knows her quite well, but Charlie thinks that Bob is a chauvinist who has a poor opinion of the mathematical abilities of women.

So we are no longer sure that $K_a(p)$, $K_bK_a(p)$ and $K_cK_bK_a(p)$ are all true.

Indeed, our confidence in our first example was misplaced. For suppose that Ann has very poor eyesight, and is currently not wearing her glasses, or perhaps Bob *thinks* that she has poor eyesight, etc.

The situation we are in is one where our common sense departs from what the Kripke semantics of knowledge tells us. Kripke semantics tells us the wrong answers, and we know they are the wrong answers, but what we need is a formal apparatus for describing real situations.

Consider now the following game. Ann is sitting (again) in a chair in front of a blackboard on which the number n is written. In front of her are three buttons, 1, 2, 3. Bob can see her and the blackboard, and we won't say yet whether she can see him. It won't matter at the start. Charlie can see both Ann and Bob and the blackboard. Bob and Charlie also have buttons. No one can see the buttons of the other people.

The game is played this way. Ann should push button 1 if she thinks n is prime, button 2 if she thinks it is composite, and button 3 if she does not know. If she presses the right button she gets one dollar. If she guesses wrong, she pays $10. And if she presses 3, there is no gain or loss. Thus the buttons stand for *prime, composite, don't know* respectively.

Bob has four buttons, and he should press a button corresponding to Ann's if he knows which button it is, and he presses button 4 if he does not know. If he guesses right, he gets $1, if he guesses wrong, he pays $10, and if he presses 4, no gain or loss.

Charlie has 5 buttons, buttons 1-4 to indicate what he thinks Bob pressed, and button 5 if he does not know. His payments are similar to Bob's.

If p denotes the fact that n is composite, then we ought to have $K_a(p), K_bK_a(p)$ and $K_cK_bK_a(p)$. Thus all three should press button 2, all of them getting[1] $1.

Will this happen? Not necessarily! As we saw, Ann may not realize that the number is composite, or if she does, Bob might think the number is too big for her to factorize etc. Thus in fact we do not have a definite map from physical situations to Kripke structures. The physical set up leaves out the mental facts, and there are many interpretations (not all of which are Kripke structures) for the *same* physical situation.

So how will the game be played? It depends, even if *some* of the three payers are logically omniscient. But note that Ann's best strategy is to press button 2 regardless of what Bob and Charlie press. *Given that she presses button 2*, Bob's best strategy is to press 2 also, and *given* that they are both pressing 2, Charlie should also play 2.

[1] But note that for Bob to be right when he presses 2, Ann has first to press button 2, and for Charlie to be right pressing 2, Bob has to have pressed 2. There is a knowledge dependence here which the Kripke structure account left out.

The *standard* Kripke structure that we get out of the physical situation does not necessarily represent the mental situation, but *it does represent the unique Nash equilibrium.*

A Formalism: We now look at a more general case, where a given physical situation is represented by a Kripke structure, but we do not assume that the knowledge of the agents can simply be read off from the Kripke structure. The latter assumes that each agent *has already* carried out the deductions which she is entitled to make, and moreover, has a right to assume that the other agents have also done so. We shall **not** make such an assumption.

Suppose we have a finite n-agent Kripke structure $\mathcal{M}$. The set of states is W with cardinality m. We use this structure to construct a game $\mathcal{G}$. Each agent is told what $\mathcal{M}$ loooks like.[2] Moreover, each agent has a set of symbols corresponding to the (finitely many) equivalence classes of that agent. I.e. the space W splits into finitely many pieces which are the equivalence classes of the agent's accessibility relation and the agent has a symbol for each such class. Thus each agent has his own alphabet. Let $[s]_i$ be i's symbol (equivalence class) for $s \in W$. Thus $s \sim_i t$ iff $[s]_i = [t]_i$. When the agent sees the symbol, he knows which equivalence class he is in, but not *where* he is in that class.

At any moment of time, some state $s \in W$ is picked with probability $1/m$. Then each agent i is given the symbol $[s]_i$. i is also given a finite set X_i of formulas with the following properties. Only atoms are negated in any formula – there are no other negations in any formula. The only connectives are $\wedge$, $\vee$, K_j, $L_j = \neg K_j \neg$. Every knowledge formula (without common knowledge) can be written in this way with all negations driven in using de Morgan's laws, etc.

If $A \wedge B$ $(A \vee B)$ is in X_i, then so are A, B.

If $K_j(A)$ or $L_j(A)$ is in X_i, then A is in X_j.

At time t, each agent i is asked to mark each formula in X_i of level $t-1$ with a *yes*, or a *no*, or a *don't know.* The process goes on until all formulas have been marked. (We could have made this a one shot game, but the extended form is a bit prettier.)

After this, each agent gets \$1 for each formula correctly marked, \$0 for each *don't know*, and is fined $\$(m \times k)$, for each incorrectly marked formula, where m is the cardinality of W, and k is the cardinality of the finite set $\bigcup X_i$. A formula marked with *don't know* is not considered marked, so we use the word "marked" for commitments where the agent is taking a risk.

A literal (atomic formula or its negation) is considered correctly marked by i iff it is true and marked *yes*, or false and marked *no.* ("true, false" are relative to $\mathcal{M}, s$.)

Formulas $A \wedge B$ and $A \vee B$ are considered correctly marked by i if the yes/no corresponds to the

[2] Is $\mathcal{M}$ common knowledge? It does not matter! Recall that we are defining knowledge from behaviour and not abstractly.

truth value at state s.

A formula $K_j(A)$ is considered to be correctly marked by i if
either $K_j(A)$ is true and A marked *yes* by j or
$K_j(A)$ is marked *no* and either A is false, or A is not marked *yes* by j.

A formula $L_j(A)$ is considered correctly marked by i if
either $L_j(A)$ is marked *yes*, it is true, and A is not marked *no* by j or
$L_j(A)$ is false, marked *no*, and A is marked *no* by j.

Formulas marked with *don't know* are not considered marked.

The important thing here is that the moves of some players are evaluated by looking at related moves of the other players. It is incorrect for Bob to say, *Ann knows that n is composite* if Ann herself has indicated that she does not know. The Kripke structure allowed no scope for Ann to enter the picture!

Each agent may have a strategy for playing this game given by the Kripke structure and the sets X_i, and we will say that an n-tuple $S = (s_1, ..., s_n)$ of strategies is *safe* for i if i does not have a negative expected value. It is *safe* if no agent makes an expected loss. Clearly the strategy where some agent says *don't know* for all formulas, is safe for him. Indeed the *don't know* strategy is safe regardless of how the other agents play. On the other hand, a strategy where an agent says *yes* for a formula A when he does not know A (i.e., where $s \models \neg K_i(A)$), can never be safe, because if he does not know A,, then there is probability at least $1/m$ that he is wrong once, and his loss of $m \times k$ will make up for all possible gains from other cases where he is accidentally right.

Definition: A *knowledge state* for n-agents is a set of *safe* strategies for them.

By contrast we might define a *belief state* to be a set of not necessarily safe strategies. Bob could have a false belief that Ann does not know that 1243 is composite. That is not (on the face of it) a false belief about *the world*, but it is a false belief nonetheless. And if Bob has such a false belief, he will make a bad move and pay for it in our game.

Theorem: The only Nash equilibrium is where each agent marks each formula correctly according to its value at s, where A is considered to be correctly marked by A if it is marked *yes* and $s \models K_i(A)$ or it is marked *no* and $s \models K_i(\neg A)$.

Proof: Straightforward by induction on formula complexity.

This theorem shows that we do not *have to be* logically omniscient, but that Nash equlibrium requires all agents to act as if they were.

Definition: Let the *knowledge depth* $d(A)$ be the maximum length of a chain of embedded knowledge operators (K or L) in formula A. We will say that a strategy s of some agent is l-complete if the agent correctly marks all formulas of knowledge depth at most l.

Lemma: Suppose all agents other than i are ℓ-complete. Then agent i can safely be $(\ell{+}1)$-complete.

Thus for agent i to infer to level $\ell + 1$ it is sufficient that other agents do infer to level ℓ. In the Ann, Bob, Charlie example, if Ann correctly infers facts (that p is true) then Bob can safely infer one level higher, and if he does, then Charlie can safely infer two levels higher.

Thus there can be evolution towards the Nash equlibrium as follows. Each agent can safely start by marking true all knowledge-free formulas which the Kripke structure says they know, and marking false all knowledge-free formulas which the Kripke structure says they know to be false. They are not dependent on other players being intelligent.

Suppose now that all the agents proceed from some level ℓ to $\ell + 1$. They are still safe since all agents were ℓ-complete. In a finite number of steps, they will arrive at a stage where all formulas A where agent i knows *whether* A according to the Kripke structure, have been marked. Now the agents have earned the maximum they possibly could and the Nash equilibrium has been reached.

We can make a stronger assertion. Starting with the strategy where all agents say *don't know* all the time, there is a sequence of changes where at each stage, *only one* agent changes his valuation of *one formula*, and which ends up with the Nash equilibrium. Moreover, no agent is unsafe at any stage of these transformations.

What about *common knowledge*? We could extend the game by saying that Ann and Bob can mark the formula $C_{a,b}(p)$ *yes*, provided it is true in the conventional sense and they *both* mark it *yes.* But now there is no individually safe way to proceed to this situation! They must do it together.

However, if the Kripke structure $\mathcal{M}$ does satisfy $C_{a,b}(p)$, then for each formula A of the form $K_aK_bK_a....K_b(p)$ (for example) there is a way for the two agents to proceed to a stage where both agents mark A with *yes.*

We can now consider the case where some agents are – or are believed to be, logically deficient by other agents. Thus suppose that of agents 1,2,3, agents 1 and 2 are logically adequate, but they know that agent 3 has no notion that other people even have minds. All three are looking at a vase of flowers. Let p stand for *There is a vase of flowers.* Then p will be common knowledge among 1 and 2, and in fact, *that 3 knows p* will be common knowledge among 1 and 2. But p cannot be common knowledge among {1,2,3}, for 3 has no notion of what 1 and 2 are thinking![3] For example, 1 cannot mark $K_3K_1(p))$ *yes*, because he cannot count on 3 marking $K_1(p)$ *yes.*

Thus there will be a sort of Nash equlibrium where agents 1, 2 are doing their best *given* 3's deficiency!

[3] Perhaps he is autistic.

This is an important fact. When we talk about Nash equlibria, we are implicitly attributing equal logical abilities to all agents, but if some agents are logically deficient, and known to be so, there can still be a 'Nash' equilibrium for the other players, which depends on their knowledge of how the deficient players have played. Spelling out 'W', 'A', 'L', 'K' in the presence of a dog is a common ploy to prevent the dog from getting too excited on hearing the word 'walk'.

Conclusion: We have defined a more general set of knowledge states than those provided by Kripke structures. Hopefully, this more flexible notion will allow us to address various puzzles like that of the *No Trade theorem*, or the issue of mathematical knowledge.

References

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

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

[3] J. Hintikka, *Knowledge and Belief*, Cornell University Press, 1962.

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

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

[6] R. Parikh, "Sentences, Propositions and Logical Omniscience, or What does Deduction tell us?", to appear in the *Review of Symbolic Logic.*