Common Knowledge

Topics in Epistemology
Oxford, 10/5/12

Daniel Rothschild
daniel.rothschild@philosophy.ox.ac.uk

1 Common Knowledge

Let's say one is WELL-INFORMED about a proposition p just in case if p is true then you know p.

Suppose we consider the set of propositions expressed by sentences formed out of only a) John's knowledge operator K_j (read $K_j s$ as *John knows* s, b) some sentence s, c) standard logical connectives. Call this set $T_{K_j,s}$

John is well informed about a set of propositions just in case John is well informed about all its members. Suppose John is well-informed about $T_{K_j,s}$. This is not so crazy: it's just saying that John knows everything there is to know about s and his knowledge of s. One consequence: For any n, $K_j^0 \ldots K_j^n s$ (where superscripts are just for numbering purposes but lack meaning). (Is there a term for this obtaining? Complete introspective knowledge about s?)

Consider generalization of this notion for situation in which more than one agents are involved. Let's suppose two and call them John and Mary, with K_j and K_m associated with each. Now suppose that John is well informed about $T_{K_j,K_m,s}$ and s is true. What follows from this? Well, not that much, since we don't know much about Mary's knowledge. What, however, if Mary is also well-informed about $T_{K_j,K_m,S}$. What follows from both their well-informedness? Well, amongst other things for any n, $K_j^0 K_n^0 \ldots K_j^n K_m^n s$.

This last statement is generally what we call COMMON KNOWLEDGE of s obtaining for the group which includes just John and Mary. (At least it's the ITERATIVE conception of it.) It seems to follow simply from mutual well-informedness about a) the underlying fact, and b) each other's knowledge states with relation to that fact and themselves.

Note also that most of the statements here known by well-informedness are also only true by well-informedness. Normally assuming well-informedness about a set of propositions does not affect their truth-values, but when these include claims about one's own knowledge, it does.

Some questions about common knowledge we may or may not answer.

- How best to characterize and reason about common knowledge.
- Its attainability.
- Its use, its implications.

2 Plausible Examples of Common Knowledge

- Visually-presented facts when jointly attended to.
- Content of credible public announcements.
- Basic knowledge (e.g. simple arithmetic, logic, common sense facts, etc)
- Rules of games (sometimes).

We might also think that common knowledge is built into the structure of certain other concepts. To take an example from Fagin et al. [1995], you might think the following principle is analytic of *agree*:

> If a group agrees that p, then each member knows that they agree that p.

Now, if we assume common knowledge of this principle (and of logical omniscience!), we can get the following:

> If a group agrees that p, then it is common knowledge that they agree that p.

Note alternative formulation of common knowledge in terms of K_e (gloss: everyone knows that…), s is common knowledge only if for any n, $K_e^1 \ldots K_e^n s$. However while such iterated knowledge claims are not the main focus of attention in epistemology generally, they play a large role in INTERACTIVE EPISTEMOLOGY, the study of these multiple agent knowledge situations. In addition, note that while the KK principle is at least plausible (i.e. that when you know something you know you know it), the $K_e K_e$ has no surface plausibility aside for special circumstances (e.g. what is known from joint attention).

Should we be skeptical of the extent of common knowledge? One reason for skepticism is worries about the $K_e K_e$ principle *even in good cases* (e.g. ones in which there is some kind of enabling condition for common knowledge).

In the old days, people seemed to worry a lot about these infinite aspect of common knowledge, because of finite minds.
Hawthorne and Magidor [2009] express skepticism, in a criticism of Stalnaker [1978], which amounts to skepticism about certain good cases actually yielding common knowledge, for essentially the arguments given by Williamson [2001] agains the KK principle.
This argument (adapted slightly) goes as follows:

MEK_e If a tree is n cm tall, then we cannot all know it is less than $n+1$ cm tall.

If we combine this principle (and everyone's knowledge of it) with the K_eK_e principle (in a given situation, due to, say, joint attention), as well as the closure of K_e we get a bad result (I think). Seems like the standard reply would be simply to deny **ME**K_e, as Stalnaker [2009] does. It looks about as plausible as a sorites premise, after all.

3 Uses

3.1 Muddy children/BBQ/Hats puzzle/Conway paradox

> Suppose there are 5 (smart) children, and four of them have muddy faces. Each child can see the other four children's faces but not their own. Suppose someone comes up to the childen and announces, "At least one of you has a muddy face. I am going to ring a bell at short intervals. If when I ring the bell you know you have a muddy face then you should clean your face then."

What happens? Answer (if the kids are idealized reasoners, etc): after the fourth ring all children with muddy faces clean them.
How to get there:
Suppose just 1 child had mud on face. Then on first ring, he would wipe his face.
Suppose 2 children had mud on their faces. Then on second ring they would wipe their faces.
Suppose 3 children had mud on their faces. Then on third ring, they would wipe their faces.
So, on fourth ring in actual situation, children will wipe their faces.

Clearly can be generalized to any situation with n children and where $m \leqslant n$ of them have mud on face.
Paradoxical element: what information was conveyed by the first utterance. If no information is conveyed, why is it necessary?
(Spell out what the information is.)

3.2 Convention

Lewis [1969] introduced the formal notion of common knowledge to explain success at a coordination problem as introduced by Schelling [1960].
Here is the simplest possible coordination game:

	A	B
A	1,1	0,0
B	0,0	1,1

Is common knowledge in any sense a necessary condition for coordination.
What follows from weaker notions?
Lewis's thought about common knowledge seems to be more one of conceptual analysis as in our case of *agree* before. He takes it as part of our concept of a convention that it be common knowledge. This separates conventions from mere regularities.

3.3 Game Theory

Implicit (and informal use) in almost all of game theory. For example prisoner's dilemma:

	C	D
C	3,3	0,4
D	4,0	1,1

Here it seems like mere rationality suffices with no common knowledge assumptions. But consider a slightly more complex game:

	C	D	O
C	3,3	0,4	.5,.5
D	4,0	1,1	0,0
O	0,0	0,0	0,10

Note that in game theory and muddy children puzzle our interest is really in iterative knowledge of a few degrees, actual common knowledge is unnecessary.

3.4 Electronic mail (paradox?)

Rubinstein [1989] gives an interesting example of the practical consequences (well, *theoretically* practical) of the failure of common knowledge. This is the quaintly titled 'electronic mail game'. This is a game in which one player has some private information that the other player does not. In particular player 1 knows which of the following two games they are playing:

With greater than $p > .5$ probability it's game I:

	A	B
A	2,2	0,-4
B	-4,0,	0,0

and

with $1 - p$ probability its game II:

	A	B
A	0,0	0,-4
B	-4,0	2,2

Now in absence of any way of communicating, it's clear that A, A will be the pair of strategies played.
Imagine, however, that a certain special kind of 'electronic' mail program is used, which sends messages and that if we are in game II above, player 1 will send a message to player 2 telling her that. There is a vanishingly small chance ϵ that the email will not arrive. Now suppose that rather than ending here (like regular email) when a message is received an 'acknowledgment' is automatically generated and sent back. And in response to an acknowledgement we also send back another acknowledgment. This cycle only stops (in a finite amount of time) when an error occurs and a message is not delivered. In the end both players get a number n on their screen representing the number of messages (either original or acknowledgments) they received before the break down.
Let us assume (as part of the structure of the game) that if player 1 knows it is game I he sends no messages. It follows that when player 1 knows it is game I he plays A (this strategy is weakly dominant). Now what if player 2 receives no messages. In this case the probability that it is game I and the message failed is:

$$\frac{\epsilon(1-p)}{\epsilon(1-p)+p} < .5$$

so his best move is to play A, unsurprisingly.
Now suppose that player I knows that it is game I, so sent a message but receives no message. Then he must reason as follows. Either the message he sent got lost, or the reply got lost. The probability his message got lost is:

$$\frac{\epsilon}{\epsilon(1-\epsilon)+\epsilon} > .5$$

So given our reasoning about what player 2 will do if he receives no message, player 1 must think that player 2 will play A with over 50% chance so player 1 should play A as well.
We have no shown that in all cases in which both player gets 0 messages the each play A. We can show by induction that if they get n messages for any n they will each play A. We proved the base step already. Now suppose that it holds that both player 1 and 2 play A when they get $n-1$ messages. Consider first player 2 if he receives n messages then he knows he sent n messages. So either his nth acknowledgment got lost (in which case player 1 only got $n-1$ acknowledgements), or its confirmation get lost. Again the probability that his message got lost is:

$$\frac{\epsilon}{\epsilon(1-\epsilon)+\epsilon} > .5$$

So with that probability player 1 got $n-1$ messages, in which case by the induction hypothesis he will play A. So player 2 will as well. Similar reasoning then applies to player 1 if he received n messages (i.e. it is more likely it is because his n+1 message got lost rather than its acknowledgment getting lost). So player 1 will play A as well if he gets n acknowledgments.
Rubinstein notes then that having common knowledge that we are in game II is very different from having large finitely iterated knowledge about it.
Note that this is *really* weird. Also note that in a system with a small ϵ and no acknowledgments things are very different!

3.5 Coordinated attack

In the distributed systems literature (whatever that is!) the following situation is sometimes considered:

> Two divisions of an army are camped on two hilltops overlooking a common valley. In the valley awaits the enemy. It is clear that if both divisions attack the enemy simultaneously they will win a battle, whereas if only one division attacks it will be defeated. The divisions do not initially have plans for launching an attack on the enemy, and the commanding general of the first division wishes to coordinate a simultaneous

attack (at some time the next day). Neither general will decide to attack unless he is sure that the other will attack with him. The generals can only communicate by means of a messenger. Normally, it takes the messenger one hour to get from one encampment to the other. However, it is possible that he will get lost in the dark or, worst yet, be captured by the enemy. Fortunately, on this particular night, everything goes smoothly. How long it will take them to coordinate an attack?

Suppose the messenger sent by general 1 makes it to general 2 with a message saying "Let's attack at dawn." Will general 2 attack? Of course not, since general 1 does not know he got the message, and thus may not attack. So general 2 sends the messenger back with an acknowledgment. Suppose the messenger makes it. Will general 1 attack? No, because now general 2 does not know he got the message, so he thinks general 1 may think that he (general 2) didn't get the original message, and thus not attack. So general 1 sends the messenger back with an acknowledgment. But of course, this is not enough either. I will leave it to the reader to convince himself that no amount of acknowledgments sent back and forth ever guarantee agreement. Note that this is true if the messenger succeeds in delivering the message every time.

Seems like what is needed (as in electronic mail paradox) is not n degrees of knowledge, but true common knowledge.

3.6 Backwards induction

Backwards induction arguments, finitely repeated prisoners dilemma. Common knowledge of rationality seems to support the backwards inductions solution. In nth round, after all, defection is only rational option. Thus it seems that in $n-1$ round it must be also, and so on. This is widely criticized as being both empirically wrong and also not intuitively what rational agents should do.

4 Partition Models

The main alternative to the iterative method of discussing common knowledge, is the partition model first suggested by Aumann [1976] (the argument of which is our topic for next week...). Assume a set of worlds W which constitute logical space. We might associated with each agent i a partition k_i over W. A partition of W is a set of disjoint sets whose union is W. Now we say that in any world w what i knows in w is just the cell of the partition k_i which w is in (which we write $k_i(w)$). More formally we can say that the proposition that i knows a proposition k , which we write $K_i(p)$ is itself just:

$$K_i(p) = \{w \in W : k_i(w) \subseteq p\}$$

Of course associated with a partition is a more familiar philosophical tool a relation r_k over W. A partition is formed by an equivalence relation, and as those familiar with modal logic know, the $S5$ axioms as modeled by Kripke semantics need an equivalence relation. So the partition model of knowledge Aumann works with essentially assumed the $S5$ axioms for knowledge. These amount to (in the most useful terms for epistemology) logical omniscience, factivity, negative introspection (negative transparency), introspection (e.g. positive transparency, the KK principle).
Using the partition method allows for an elegant definition of common knowledge (Aumann's). In particular the partition associated with what is common knowledge is just the finest coarsening of all the partitions of the group (the meet). (A coarsening of a partition is just another partition that can only contain unions of cells of the first partition as its cells.) Aumann shows how (in the $S5$ model) this condition is equivalent to the iterative one.

5 Situational conception

Barwise [1988] attributes this to Clark and Marshall but it seems to be exactly what Lewis has in mind in *Convention* (as a way of generating the iterative conception).
A variation on Lewis's notion is as follows. That the state of affairs A obtains is common knowledge if:

1. You and I both know that A obtains.

2. A obtaining entails (or something else?) that you and I know that A obtains.

More generally you might think that a fact p is common knowledge iff there is a state of affair A such that:

1. You and I both know that A obtains.

2. A obtaining entails (or something else?) that you and I know that A obtains.

3. A obtaining entails (or something else?) that p.

Note that situational conception only gives you the iterative conception with closure assumptions. Barwise [1988] argues that the iterative conception is far too strong to be useful as a notion of *knowledge* rather than *information*. He writes, "Information travels at the speed of logic, genuine knowledge only travels at the speed of cognition and inference." (This seems perhaps wrong for various forms of implicit knowledge.)
We might need to be liberal in what counts as states of affairs if we want to cover all common knowledge.

6 Common Information

The notion of common knowledge given by the partitional model might be seen as a sort of notion of common information. For the notion "x has the information that p" the partitional model seems appropriate: x has the information that p might imply that x has the information that x has the information that p, etc. Certainly logical entailment holds.
Barwise writes:

> Much of the work in logic which seems to about knowledge is best understood in terms of having information. And for good reason. For examples in dealing with computers, there is a good reason for our interest in the latter notion. We often use computers as information processors, after all, for our own ends. We are often less interested in what the computer does with the information it has, than in just what information it has and what we can do with it.

7 Distributed Knowledge

Distributed knowledge is the logical consequences of all the knowledge of a group. It is what a group knows together. In the partition framework it is the *join* (the coarsest common refinement of the partitions) of the knowledge-partitions of the different members of the group.

References

Robert J. Aumann. Agreeing to disagree. *Annals of Statistics*, 4:1236–1239, 1976.

John Barwise. Three views of common knowledge. In Moshe Y. Vardi, editor, *Proceedings of TARK 2*, 1988.

Ronald Fagin, Joseph Y. Halpern, Yoram Moses, and Moshe Y. Vardi. *Reasoning about Knowledge*. MIT, 1995.

John Hawthorne and Ofra Magidor. Assertion, context, and epistemic accessibility. *Mind*, 118:377–397, 2009.

David Lewis. *Convention*. Harvard University Press, 1969.

Ariel Rubinstein. The electronic mail game: Strategic behavior under "almost common knowledge". *The American Economic Review*, 79:386–391, 1989.

Thomas C. Schelling. *The Strategy of Conflict*. Harvard University Press, 1960.

Robert Stalnaker. Assertion. In P. Cole, editor, *Syntax and Semantics*. Academic Press, 1978.

Robert Stalnaker. On Hawthorne and Magidor on assertion, context, and epistemic accessibility. *Mind*, 118:399–409, 2009.

Timothy Williamson. *Knowledge and its Limits*. OUP, 2001.