Common Knowledge and Common Belief

Jan van Eijck, CWI Amsterdam and Uil-OTS Utrecht
Rineke Verbrugge, Institute of AI, University of Groningen

ESSLLI 2009, Bordeaux, July 23, 2009

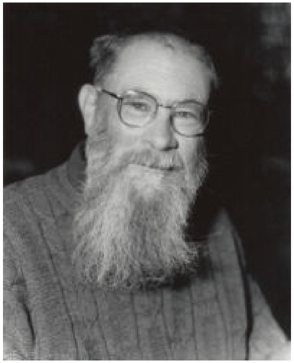
Overview

- Conventions and Common Knowledge
- Common Belief
- Achieving Common Knowledge
  - Co-presence Creates Common Knowledge
  - When Message Exchance Cannot Create Common Knowledge
- Common Knowledge and the Public Arena
- Defining Common Knowledge by Co-Recursion
- Common Belief and the Stock Market
- Agreeing to Disagree
- Defining Common Knowledge by Fixpoints

- Example: Card Distributions
- Public Announcements and Common Knowledge
- The Muddy Children

Conventions and Common Knowledge

Philosopher: Today, I suggest we discuss the important concepts of common knowledge and common belief. One of David Lewis' examples in Convention [8] is traffic conventions.



Conventions and Common Knowledge

Philosopher: Imagine driving on a one-lane road. You just came out of the Channel Tunnel on the British side. Drivers who just went from France to England tend to forget on which side of the road they have to drive, particularly if they find themselves on a quiet one-lane road where they are suddenly confronted with oncoming traffic. Then you will have to swerve a bit to the side to let it pass. In fact, you each have to swerve a bit.

Will you swerve left or right?

Cognitive Scientist: If I remember that I am in England, I will swerve left. Otherwise, I will swerve right.

Philosopher: Yes, and how about the guy coming towards you? He and you may both be cautious drivers, but if he swerves right and you left, you will still crash.

Conventions and Common Knowledge

Philosopher: It is not enough for you and the on-comer both to know that you have to drive left. You would also like to know that the other knows. And this will affect your behavior.

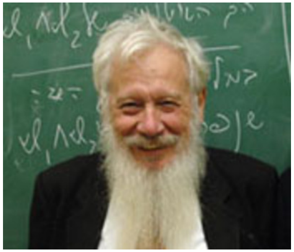
You are very cautious if you do not know, slightly less cautious if you know but not if the other knows, even less cautious if you know and also know that the other knows but not if he knows that, and so on: you become more and more confident about the other's road behavior but never entirely so.

Driving on the left-hand side is what Lewis calls a convention, and this means that everyone knows that everyone knows know that everyone knows. . . up to any finite stack of knowledge operators.

Logician: Exactly, that's common knowledge.

Common Belief

Common belief means that everyone believes that everyone believes that . . . and so on.



Economist: It is commonly believed among economists that Aumann was the first to give a formal analysis of common knowledge [1].

Philosopher: And it is commonly believed among philosophers that Lewis was the first [8].

Common Belief

Logician: Dov Samet told me that those common beliefs were wrong: Friedell was earlier than both Lewis and Aumann [4].

Philosopher: Maybe it was even commonly believed among economists that it was common knowledge that Aumann was the first to give this formal analysis. And that common belief was also false.

Achieving Common Knowledge

Computer Scientist: Common knowledge is often easily achieved, by means of public announcement.

Cognitive Scientist: And what do you mean by public announcement, exactly?

Computer Scientist: A public announcement is an event where something is being said aloud, while everybody is aware of who is present, and it is already common knowledge that all present are awake and aware, and that everybody hears the announcement, and that everybody is aware of the fact that everybody hears it, and . . .

Economist: Are e-mail notifications proper public announcements?

Message Exchange Cannot Create Common Knowledge

Logician: It was proved by Halpern and Moses [5] that message exchange in a distributed environment, where there is no guarantee that messages get delivered, cannot create common knowledge.

Computer Scientist: Analysis of message passing through unreliable channels is old hat in computer science. We call it the coordinated attack problem .

Two generals are planning a coordinated attack on a city. They are on two hills on opposite sides of the city, each with their own army, and they know they can only succeed in capturing the city if their two armies attack at the same time. But the valley that separates the two hills is in enemy hands, and any messengers that are sent from one army base to the other run a severe risk to get captured. The generals have agreed on a joint attack, but they still have to settle the time.

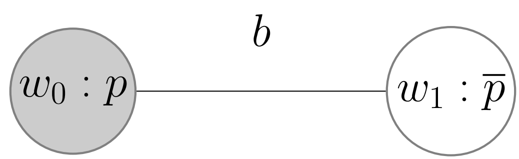
picture by
Marco Swaen

Message Exchange Cannot Create Common Knowledge

Computer Scientist: You got the picture.

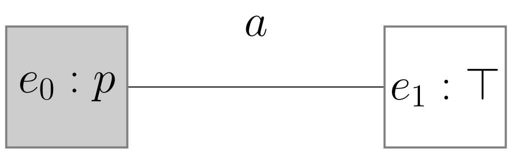
Philosopher: So the generals start sending messengers. But they cannot be sure that the messengers succeed in delivering their message. And if they get through, there is no guarantee that the message of acknowledgement will get delivered. And so on.

Two generals a, b. a will attack at 8 AM (p), but b does not know this:

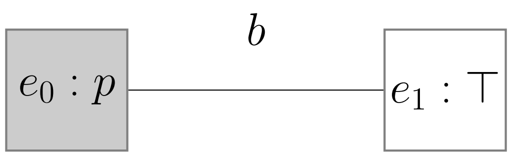


Message Exchange Cannot Create Common Knowledge

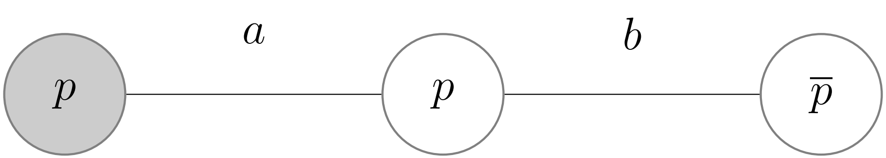
Update action for general a: send a message p.



Update action for general b: send an acknowledgement of p:

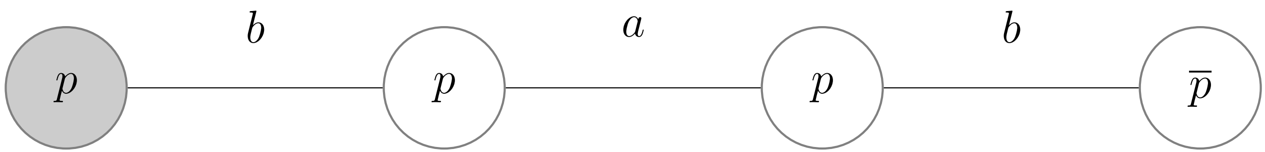


Situation after first message from general a:



Message Exchange Cannot Create Common Knowledge

Situation after update by a followed by update by b:



And so on . . .

Computer Scientist: The procedure is known in computer science as the "alternating bit protocol" for sending bits over an unreliable channel. The sender repeats the transmission of a bit until an acknowledgement is received, then the sender acknowledges the receiver's acknowledgment until that is in turn acknowledged by the receiver, and only then the next bit is sent until that bit gets acknowledged, and so on. Still, this will not achieve common knowledge .

Co-presence Creates Common Knowledge

Philosopher: What are the properties of events that succeed in creating common knowledge? It seems to me that they all involve a shared awareness that a common experience takes place. It can involve various senses: hearing, eye-contact, maybe even touching or smelling.

> If B sees A look at B, then A sees B look at A. From this and a few simpler properties one can demonstrate that eye contact leads to common knowledge of the presence of the interactants. It is no coincidence that eye contact is of considerable emotional and normative significance (Friedell, [4, page 34]).

Co-presence Creates Common Knowledge

Example: cash withdrawal from a bank.

You withdraw a large amount of money from your bank account and have it paid out to you in cash by the cashier.

The cashier looks at you earnestly to make sure she has your full attention, and then she slowly counts out the banknotes for you: one thousand (counting ten notes), two thousand (counting another ten notes), three thousand (ten notes again), and four thousand (another ten notes).

This ritual creates common knowledge that forty banknotes of a hundred euros were paid out to you.

Philosophical question: when money is paid out to you by an ATM, does this create common knowledge between you and the machine?

Common Knowledge and the Public Arena

Cognitive Scientist: Yes, but how does one know that an announcement has become common knowledge? I might have let my attention wander for a moment.

Computer Scientist: If an announcement is made, you were supposed to pay attention, and therefore the information can now be assumed common knowledge.

Philosopher: That is what happens in the public arena all the time. At the basis of legal relations between individuals and the state, is the assumption that the law is common knowledge.

Cognitive Scientist: But this is a fiction. Professional lawyers have a full-time job to keep up with the law. Ordinary citizens can simply not be expected to cope.

Common Knowledge and the Public Arena

Philosopher: I prefer to say that it is a necessary presumption. Roman lawgivers found out long ago that if citizens within their jurisdiction could plead innocence because of being unaware of the law, no offender could ever get convicted. So they were quick to invent principles like Ignorantia legis neminem excusat, "ignorance of the law excuses no one".

Computer Scientist: As a counterpart the laws have to be properly published and distributed. Of course, the citizens are not supposed to read all that boring stuff, but they should be able to find out about it whenever they want. In this way, the publications in the government gazette amount to public announcements.

Olmert's Nuclear Slip-up

Olmert's Nuclear Slip-up

Ehud Olmert, the Israeli Prime Minister, faced calls for his resignation today after admitting - in an apparent slip of the tongue - that Israel has got nuclear weapons.

But Israeli officials tried to push the cat back into the bag, denying that Mr Olmert had made any such admission and falling back on the Jewish state's policy of "nuclear ambiguity".

Widely considered the Middle East's sole nuclear power, Israel has for decades refused to confirm or deny whether it possesses the atomic bomb. Mr Olmert appeared to break that taboo in an interview with a German television station as he began a visit to Berlin.

TimesOnline, Dec 12, 2006

Olmert's Nuclear Slip-up

Challenge for Social Software Analysis: design a framework in which "public φ-hypocrisy" can be formally expressed and analyzed.

Defining Common Knowledge by Co-recursion

General knowledge $E\varphi$ for set of agents $\{1, \ldots, n\}$:

$E\varphi \leftrightarrow K_1\varphi \wedge \ldots \wedge K_n\varphi$

Logician: General knowledge among the members of a group of agents means that all individuals in the group know a certain fact, and common knowledge means: everybody knows that everybody knows, and so on.

Computer Scientist: Let me propose a definition of common knowledge: A proposition φ is common knowledge if everybody knows that φ and everybody knows that φ is common knowledge:

$C\varphi \leftrightarrow E\varphi \wedge EC\varphi$

Philosopher: That can hardly qualify as a definition, it's obviously circular.

Defining Common Knowledge by Co-recursion

Computer Scientist: "$C\varphi$ iff $E\varphi \wedge EC\varphi$" is an instance of a definition by co-recursion. They are like recursive definitions, but with the crucial difference that there is no base case. And they define infinite objects.

An infinite stream of zeros, call it zeros, can be defined as: zeros equals a zero followed by zeros. In lazy functional programming this is written as

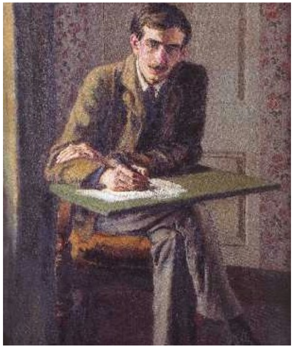
$$\text{zeros} = 0 : \text{zeros}$$

If you execute this program in Haskell you will get an infinite stream of zeros flashing over your screen.

Common Belief and the Stock Market

[..] professional investment may be likened to those newspaper competitions in which the competitors have to pick out the six prettiest faces from a hundred photographs, the prize being awarded to the competitor whose choice most nearly corresponds to the average preferences of the competitors as a whole; so that each competitor has to pick, not those faces which he himself finds prettiest, but those which he thinks likeliest to catch the fancy of the other competitors, all of whom are looking at the problem from the same point of view. It is not a case of choosing those which, to the best of one's judgment, are really the prettiest, nor even those which average opinion genuinely thinks the prettiest. We have reached the third degree where we devote our intelligences to anticipating what average opinion expects the average opinion to be. And there are some, I believe, who practise the fourth, fifth and higher degrees [6]

Common Belief and the Stock Market



Philosopher: The quote from Keynes talks about levels of belief, and in the limit about common belief. The prize in the beauty contest does not go to the person who picks the prettiest girl, but to the person who picks the girl that is commonly believed to be the prettiest girl. If Keynes is right that the stock-market is about common belief, then as long as a stock is commonly believed to be worth a lot, it does not matter if it is overvalued.

Common Belief and the Stock Market

Logician: That reminds me of the current credit crunch. Imagine a rumor that a bank is going to go bankrupt. The rumor may be false, but it can start a chain reaction which results in the bank actually going bankrupt. If we want to be serious about social software, we need to be able to explain such a phenomenon, and possibly even to devise mechanisms to prevent them.

Economist: Behavioral economists Bicchieri and Xiao [2] investigate how social norms influence individual decision making.

Common Belief and the Stock Market

Economist: It turns out that what we expect others to do significantly predicts our own choices, much more than what we expect others to think we ought to do. Such findings are important if you want to design policies aimed at discouraging undesirable behavior.

Logician: That's what I always tell my spouse: It doesn't help to tell our children not to smoke or drink or lie: We should consistently set the right example...

Agreeing to Disagree

Economist: In the economics setting, instead of different possible situations—such as driving on the left, or on the right—the preferred model is that of different probable situations, and how events relate prior to posterior probabilities. In "Agreeing to disagree" [1], Aumann shows that if agents have common knowledge of their posterior probabilities of an event, that these must then be the same. In other words, they can only agree to agree and they cannot agree to disagree.

It is not rational to agree to disagree, because this agreement would entail awareness of the fact that the disagreement can be exploited.

Agreeing to Disagree

Logician: What does it mean that you believe that the probability of an event is $\frac{1}{2}$? Simply that if you are taking bets on this, then you will consider a bet with a return of two to one a fair bet. And if you believe that the probability is $\frac{1}{4}$ and you are in a betting mood, then you will consider a bet with a return of four to one (including the stake) a fair bet.

Computer Scientist: That's what bookies call an odds of three to one against: If the event happens you win three times your stake, otherwise you lose your stake.

A Dutch Book about ESSLLI 2011?

Will ESSLLI be in St Petersburg or in Salamanca?

Jan: probability of ESSLLI in St Petersburg $\frac{1}{2}$
Jan is willing to take odds of one to one against St. Petersburg.

Rineke: probability of ESSLLI in St Petersburg $\frac{1}{4}$
Rineke is willing to take odds of three to one against St. Petersburg.

A Dutch Book about ESSLLI 2011?

Computer Scientist: Assume a student places two bets:

- A bet of 1000,- to Rineke that St Petersburg will win (for three to one against).
- A bet of 2000,- to Jan that Salamanca will win (for one to one against).

If St. Petersburg wins, the student collects 3000,- from Rineke and loses her 2000,- to Jan: gain of 1,000.

If Salamanca wins, the students loses her stake of 1,000 to Rineke but collects 2,000 from Jan: gain of 1,000.

A Dutch book is a set of odds and bets which guarantees a profit. Agreeing to disagree is not rational for people willing to take bets on their beliefs.

Defining Common Knowledge by Fixpoints

Philosopher: Let us move on to the logic of common knowledge. How do we know that the concept of common knowledge is well-defined? And how do we know that common knowledge can be achieved in a finite number of steps?

Logician: The answer to the first question lies in a famous theorem by Tarski and Knaster. Let F be the operation of mapping a set of situations X to the set of situations where X is general knowledge and where $F(X)$ is also general knowledge. Then this operation is monotonic. This means that it preserves the ordering on situations. If X is less informative than Y then $F(X)$ will also be less informative then $F(Y)$. Then F is guaranteed to have a fixpoint.

Philosopher: What do you mean by "less informative"? And what is a fixpoint?

Defining Common Knowledge by Fixpoints

Logician: For sets of situations "less informative" will be reverse inclusion: If you can exclude more situations, you know more. Tarski and Knaster [7] prove that all monotonic functions have fixpoints. A fixpoint of a function F is a value X for which $F(X) = X$.

Logician: An agent a is said to know φ in a state s if the proposition φ holds in all states that a cannot distinguish from s.

Two agents, Alice and Bob, want to achieve common knowledge on who is going to collect the kids from daycare.

$$1 — \text{Alice} — 2 — \text{Bob} — 3 — \text{Alice} — 4 — \text{Bob} — 5 — \cdots$$

Suppose a path links state 1 to state 2 for Alice, followed by a link from 2 to 3 for Bob, and so on. Something is common knowledge for Alice and Bob if it is true in all situations on such a path.

Defining Common Knowledge by Fixpoints

Philosopher: Ah, now I see how fixpoints come in. For common knowledge you have to compute the transitive closure of the union of the accessibility relations for Alice and Bob.

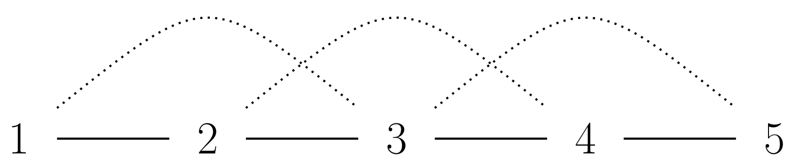
Computer Scientist: The fixpoint procedure for making a relation transitive goes like this:

1. Check if all two-step transitions can be done in a single step. If so, the relation is transitive, and done.
2. If not, add all two-step transitions as new links, and go back to 1.

Defining Common Knowledge by Fixpoints

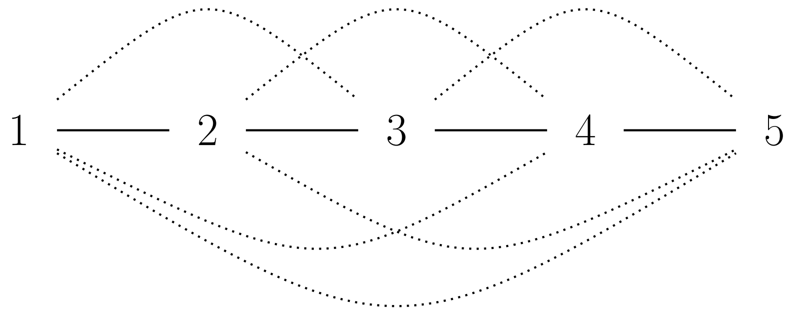
Philosopher: I suppose we can think of the link from 1 to 2 as a link for Alice, and the link from 2 to 3 as a link for Bob, and so on?

Computer Scientist: That's right, but I have blurred the distinction by taking the union of Alice's and Bob's links. Our check reveals that not all two step transitions can be done in single leaps, so the relation is not transitive. In the first step, we add all two-step links as new links:



Defining Common Knowledge by Fixpoints

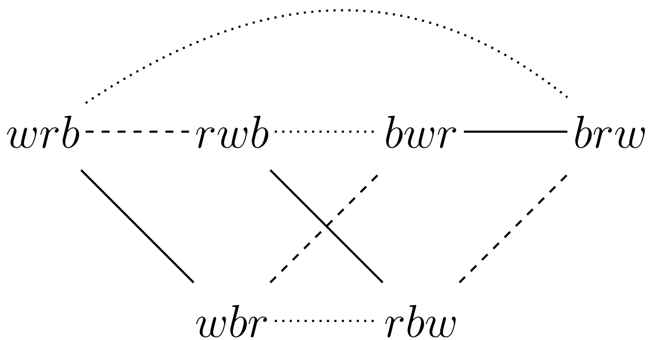
Computer Scientist: Now we check again. No, this is not yet transitive. So we add all two-step links in this new picture as extra links:



Philosopher: This is indeed a fixpoint procedure. You change the relation step by step, until it has the required property. After your final step the relation has become transitive. So a proposition is common knowledge between Alice and Bob if it is true in all those states.

Example: Card Distributions

Computer Scientist: Consider the situation where Alice, Bob and Carol each receive a card from the set red, white and blue. They can all see their own card, but not those of the others [3]:



wbr represents the state in which Alice has white, Bob has blue and Carol has red, etc. The solid lines are for Alice. If she has white, she can see that she has white, but she cannot distinguish wbr from wrb.

Philosopher: Let me see. Then the dotted arrows must represent Bob's knowledge relation, and the dashed arrows Carol's. So now one can say things like "Alice holds white" by means of propositional atoms such as w_{Alice}.

Philosopher: So in situation wbr it is common knowledge among Alice, Bob and Carol that Alice doesn't know that Bob has blue?

$$wbr \models C_{\{Alice,Bob,Carol\}} \neg K_{Alice} b_{Bob}$$

Computer Scientist: That's right. This is because all six worlds can be reached from wbr in one or more steps by accessibility relations for agents in the group, and it is clear that $\neg K_{Alice} b_{Bob}$ in all worlds, for it holds everywhere that Alice can access at least one world in which Bob doesn't have blue.

Public Announcements and Common Knowledge

Logician: Suppose Alice suddenly makes the public announcement: “I am holding white”. Then we can remove all other worlds where she does not hold white from the model:

$$wrb \text{———} wbr$$

Now it has become common knowledge that Alice holds white. And it is common knowledge that the only uncertainty that remains is Alice's uncertainty about the cards of Bob and Carol.

Cognitive Scientist: So, does publicly announcing φ always lead to common knowledge of φ?

Public Announcements and Common Knowledge

Logician: Well, not quite. Suppose instead of “I am holding white”, Alice would have announced “I am holding white, but you guys don’t know it yet.” Then the second part of this becomes false as an effect of the announcement.

Philosopher: Alice is using a variation on the famous Moore sentence [9, p.543]: “I went to the pictures last Tuesday, but I don’t believe that I did.”

Cognitive Scientist: Moore sentences have the property that you cannot truthfully repeat them. So indeed, not all φ can be made common knowledge by publicly announcing them.

The Muddy Children

Logician: Among n children, there are k (which is at least one) of them with mud on their foreheads. They can see each other but not themselves. Now their father confronts them and says aloud: 'At least one of you has mud on his forehead. Will all the children who know they have mud on their heads please step forward?'

picture by
Marco Swaen

The Muddy Children

First, none of the children step forward. When the father repeats his question, he will still get no response until he asks the question for the k-th time. Then, miraculously, all muddy children step forward.

Solution to Homework: Protocol for Prisoners with Lightbulb

Assume there are $n > 2$ prisoners.

The n prisoners appoint one among them as the 'counter'.

All prisoners except the counter act as follows: the first time they enter the room when the light is off, they switch it on; on all next occasions, they do nothing.

The counter acts as follows: The first $n - 2$ times that the light is on when he enters the interrogation room, he turns it off. Then the next time he enters the room when the light is on, he (truthfully) announces that everybody has been interrogated.

Analysis

- Last year's ESSLLI course gave an analysis in terms of dynamic epistemic logic. See:

  `http://www.cwi.nl/~jve/courses/esslli08`

- How would you state the correctness property of the protocol?
- Why should the prisoners ever agree on carrying out the protocol?
- How do they all know that at some point in the future the counter will know that they have all been interrogated?

References

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

[2] C. Bicchieri and E. Xiao. Do the right thing: But only if others do so. Journal of Behavioral Decision Making, 21:1–18, 2008.

[3] H.P. van Ditmarsch, W. van der Hoek, and B.P. Kooi. Dynamic Epistemic Logic, volume 337 of Synthese Library. Springer, 2007.

[4] M.F. Friedell. On the structure of shared awareness. Behavioral Science, 14(1):28–39, 1969.

[5] J.Y. Halpern and Y. Moses. Knowledge and common knowledge in a distributed environment. In Proceedings of the 3rd ACM Symposium on Principles of Distributed Computing (PODS), pages 50–61, 1984.

A newer version appeared in the Journal of the ACM, vol. 37:3, 1990, pp. 549–587.

[6] John Maynard Keynes. The General Theory of Employment, Interest and Money. Macmillan and Cambridge University Press, 1936. Full text available on the internet at http://www.marxists.org/reference/subject/economics/keynes/general-theory/.

[7] B. Knaster. Un théorème sur les fonctions d'ensembles. Ann. Soc. Polon. Math, 6:133–134, 1928.

[8] D.K. Lewis. Convention: A Philosophical Study. Harvard University Press, Cambridge (MA), 1969.

[9] G.E. Moore. A reply to my critics. In P.A. Schilpp, editor, The Philosophy of G.E. Moore, pages 535–677. Northwestern University, Evanston IL, 1942. The Library of Living Philosophers (volume 4).