Dealing with Other Agents: Communication and Common Knowledge

Alan Bundy
School of informatics
University of Edinburgh

(some slides from Joe Halpern)

Nested Knowledge

K_A K_B φ: reasoning about other's knowledge.

K_A K_A φ: reasoning about your own knowledge.

K_A K_B K_A φ: reasoning about other's knowledge of you.

K_A K_B K_C φ: reasoning about other's knowledge of others.

Applications of K_A φ

Reasoning about collaborators: multi-agent planning,

e.g. for organising disaster recovery operations.

Reasoning about opponents: in games, in business, in warfare, *etc*,

e.g. for softbot web negotiating and trading.

Reasoning about yourself: reflection,

e.g. leading to belief revision.

The Muddy Children Puzzle

A number, say n, of children are standing in a circle around their father. There are k ($1 \leq k \leq n$) children with mud on their heads. The children can see each other but they cannot see themselves. In particular they do not know if they themselves have mud on their heads. There is no communication between the children. The children all attended a course on epistemic logic and they can reason with this in a perfect way. Furthermore, they are perfectly honest and do not cheat. Now Father says aloud: 'There is at least one child with mud on its head. Will all the children who know they have mud on their heads please step forward?' In case $k > 1$, no child steps forward. Father repeats his question. If $k > 2$, again the children show no response. This procedure is repeated until, after the k^{th} time Father has asked the same question, all k muddy children miraculously step forward.

Why?

Muddy Children Puzzle

How to Understand Muddy Children Puzzle

- What does a clean child see? k dirty children.
- What does a dirty child see? $k-1$ dirty children.
- Suppose $k=1$: dirty child sees 0 dirty children,
  so knows immediately s/he is dirty.
- Suppose $k=2$: both dirty children see 1 dirty child,
  so neither steps forward on first iteration.
  When they see this they realise they must be dirty,
  So step forward on second iteration.
- In general, after k iterations, all dirty children realise they are dirty.

Possible World Account

- Let $\langle b_1, \ldots, b_n\rangle$ be possible world,
  where $b_i = 1$ iff child i is muddy and $b_i = 0$ otherwise.
- For concreteness let $n = 3$, $k = 2$ and $\langle 1, 0, 1\rangle$ be actual world.
- What worlds will each child consider possible?
  - Child 1 will consider $\langle 1, 0, 1\rangle$ and $\langle 0, 0, 1\rangle$.
  - Child 2 will consider $\langle 1, 0, 1\rangle$ and $\langle 1, 1, 1\rangle$.
  - Child 3 will consider $\langle 1, 0, 1\rangle$ and $\langle 1, 0, 0\rangle$.

Eliminating Possible Worlds

- What does child 1 know about child 3's possible worlds?
- Child 3 may consider: $\langle 1, 0, 1\rangle$, $\langle 1, 0, 0\rangle$, $\langle 0, 0, 1\rangle$ and $\langle 0, 0, 0\rangle$.
- But when Father says $k \geq 1$ then $\langle 0, 0, 0\rangle$ is impossible.
- Since child 3 does not then step forward then $\langle 0, 0, 1\rangle$ also impossible.
- So child 1 knows child 3 is considering $\langle 1, 0, 1\rangle$ and $\langle 1, 0, 0\rangle$,
  in both of which child 1 is dirty.
- So child 1 steps forward on second iteration.

Mid-Lecture Exercise

In the muddy children puzzle, let a possible world be denoted by $\langle b_1, \ldots, b_n \rangle$, where $b_i = 1$ iff child i has mud on his/her head. Suppose $n = 4$ and the actual world is $\langle 1, 0, 1, 0 \rangle$. Which of the following worlds will child 2 consider to be possible?

1. $\langle 1, 1, 1, 0 \rangle$ and $\langle 1, 0, 1, 1 \rangle$.
2. $\langle 1, 0, 1, 0 \rangle$ and $\langle 1, 0, 1, 1 \rangle$.
3. $\langle 0, 0, 1, 0 \rangle$ and $\langle 1, 0, 1, 0 \rangle$.
4. $\langle 1, 0, 1, 0 \rangle$ and $\langle 1, 0, 0, 0 \rangle$.
5. $\langle 1, 0, 1, 0 \rangle$ and $\langle 1, 1, 1, 0 \rangle$.

Solution to Exercise

Child 2 will only consider the following two worlds to be possible:

5. $\langle 1, 0, 1, 0 \rangle$ and $\langle 1, 1, 1, 0 \rangle$.

since the *actual* world is $\langle 1, 0, 1, 0 \rangle$ and child 2 can only be in doubt about whether mud is on his/her head.

Role of Father's Statement

- We can prove by induction on k that if k children have muddy foreheads, they say "yes" on the k^{th} question.
- It appears as if the father didn't tell the children anything they didn't already know.
- Yet, without the father's statement, they could not have deduced anything.
- So what was the role of the father's statement?

States of Group Knowledge

When reasoning about the knowledge of a group of agents, states of "group knowledge" become relevant:

- someone knows
- everyone knows
- everyone knows that everyone knows
- everyone knows that everyone knows that everyone knows

  ...
- common knowledge

We have a hierarchy:

$$\boxed{\mathsf{C}}\,\varphi \Rightarrow \ldots \Rightarrow \boxed{\mathsf{E}}\,\boxed{\mathsf{E}}\,\boxed{\mathsf{E}}\,\varphi \Rightarrow \boxed{\mathsf{E}}\,\boxed{\mathsf{E}}\,\varphi \Rightarrow \boxed{\mathsf{E}}\,\varphi \Rightarrow \boxed{\mathsf{K}_A}\,\varphi$$

Communication moves you up the hierarchy.

Common Knowledge

$\boxed{\mathsf{E}}\,\varphi$ **everybody knows** that φ

$\boxed{\mathsf{C}}\,\varphi$ it is **common knowledge** that φ

where

$$\boxed{\mathsf{E}}\,\varphi \Leftrightarrow \boxed{\mathsf{K}_A}\,\varphi \wedge \boxed{\mathsf{K}_B}\,\varphi \wedge \boxed{\mathsf{K}_C}\,\varphi \wedge \ldots$$

$$\boxed{\mathsf{C}}\,\varphi \Leftrightarrow \boxed{\mathsf{E}}\,\varphi \wedge \boxed{\mathsf{E}}\,\boxed{\mathsf{E}}\,\varphi \wedge \boxed{\mathsf{E}}\,\boxed{\mathsf{E}}\,\boxed{\mathsf{E}}\,\varphi \wedge \ldots$$

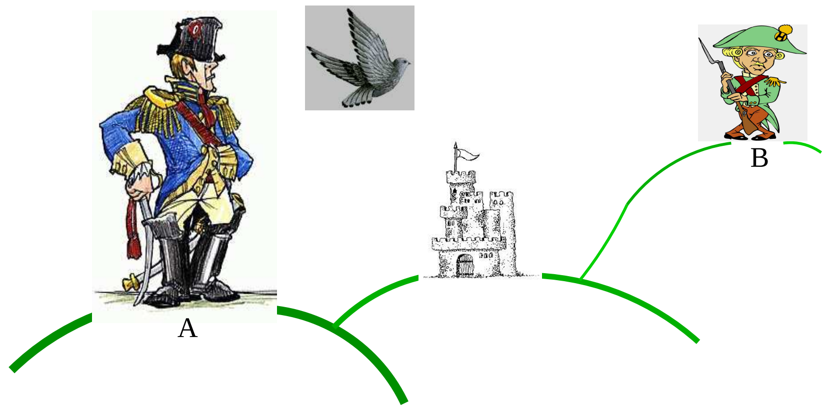
or

$$\boxed{\mathsf{C}}\,\varphi \Leftrightarrow \boxed{\mathsf{E}}\,(\varphi \wedge \boxed{\mathsf{C}}\,\varphi)$$

Acquiring Common Knowledge

- Deadlock detection algorithms convert a situation where the group has distributed knowledge of the deadlock to one where everyone knows about it (and so can take appropriate action).
- Communication conventions must be common knowledge.
- Agreement requires common knowledge.
- The father gives the children *common knowledge* of the fact that at least one child has a muddy forehead.

Attack at Dawn



The Coordinated Attack Problem

- Each time the messenger makes it, the level of knowledge rises.
- Let m = "General A sent a message saying 'attack at dawn'."
- First $\boxed{\mathsf{K}_B}\,m$, then $\boxed{\mathsf{K}_A}\,\boxed{\mathsf{K}_B}\,m$, $\boxed{\mathsf{K}_B}\,\boxed{\mathsf{K}_A}\,\boxed{\mathsf{K}_B}\,m$, ...
- **Proposition:** (Halpern-Moses) m will never become common knowledge using a k-round handshake protocol.
- **Theorem:** m will never become common knowledge in any run of any protocol. In fact, common knowledge is not attainable in any system where communication is not guaranteed.

Impossibility of Coordinated Attack

- But what about coordinated attack?
- *Agreement implies common knowledge.*
- **Corollary:** Any protocol that guarantees that if one of the generals attacks, then the other does so at the same time, is a protocol where necessarily neither general attacks.

  (N.B. We need to also assume that in the absence of messages, neither general will attack.)

Variants on this Theme

- We have shown that common knowledge is not attainable if communication is not guaranteed.
- We can easily show that common knowledge is also not attainable if communication is guaranteed, but there is no upper bound on message delivery time.
- Even if there *is* an upper bound on message delivery time, but the actual message delivery time is uncertain, common knowledge is not attainable.

Why Upper Bound not Sufficient

- Agents A and B are trying to attain $\boxed{\mathsf{C}}\varphi$.
- Suppose communication guaranteed within either 0 or ϵ seconds.
- At time t_A, A sends φ which is received by B at t_B.

  Note either $t_B = t_A$ or $t_B = t_A + \epsilon$.
- $\boxed{\mathsf{K}_A}\boxed{\mathsf{K}_B}\varphi$ only at time $t_A + \epsilon$, in case $t_B = t_A + \epsilon$
- $\boxed{\mathsf{K}_B}\boxed{\mathsf{K}_A}\boxed{\mathsf{K}_B}\varphi$ only at time $t_B + \epsilon$, in case $t_B = t_A$.
- $\boxed{\mathsf{K}_A}\boxed{\mathsf{K}_B}\boxed{\mathsf{K}_A}\boxed{\mathsf{K}_B}\varphi$ only at time $t_A + 2\epsilon$.
- $(\boxed{\mathsf{K}_A}\boxed{\mathsf{K}_B})^k\varphi$ only at time $t_A + k\epsilon$.
- So common knowledge never attained.

Solution: Synchronised Clocks

- The situation is very different if there is a global clock and the messages are timestamped:
- If A says "m; the time is 5am", this message becomes common knowledge at $5 + \epsilon$.

  since B knows which of $t_B = t_A$ or $t_B = t_A + \epsilon$ is true.
- **Theorem:** Common knowledge requires synchronized clocks.
- **Corollary:** In any system where message delivery time is uncertain and clocks are not initially synchronized, common knowledge is not attainable.

Conclusion

- Common knowledge necessary when coordinated action required.
- Need protocols that ensure common knowledge.
- k-round handshakes cannot attain common knowledge,

  cf. coordinated attack problem.

  even when communication guaranteed.
- Common knowledge requires synchronised clocks.

  e.g. by public announcement, *cf. muddy children puzzle.*