Alternative axiomatisations of common knowledge

Andreas Herzig, Elise Perrotin
CNRS, IRIT, France

Rennes, FMAI Workshop, May 2, 2019

Epistemic language

$$\varphi \ ::= \ p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \mathbf{K}_i\varphi \mid \mathbf{E}\varphi \mid \mathbf{C}\varphi$$

(only one, fixed group)

- $\mathbf{K}_i\varphi$ = "agent i knows that φ"
- $\mathbf{E}\varphi = \bigwedge_{i \in Agt} \mathbf{K}_i\varphi$ = "it is shared knowledge that φ"
- $\mathbf{C}\varphi = \bigwedge_{k \geq 0} \mathbf{E}^k\varphi$ = "it is common knowledge that φ"

S5 individual knowledge

$$\boxed{\mathrm{S5}(\mathbf{K}) \;=\; \text{modal logic S5 for the modal operators } \mathbf{K}_i}$$

- truth axiom:

$$\mathbf{K}_i\varphi \to \varphi$$

- positive introspection axiom:

$$\mathbf{K}_i\varphi \to \mathbf{K}_i\mathbf{K}_i\varphi$$

- negative introspection axiom:

$$\neg\mathbf{K}_i\varphi \to \mathbf{K}_i\neg\mathbf{K}_i\varphi$$

Shared knowledge: definition

$$\boxed{\texttt{Def}(\mathbf{E})\text{: } \mathbf{E}\varphi \leftrightarrow \bigwedge_{i \in Agt} \mathbf{K}_i\varphi}$$

- normal modal operator:
  - axiom $\texttt{K}(\mathbf{E})$ provable
  - rule of necessitation $\texttt{RN}(\mathbf{E})$ derivable
- truth axiom provable:

$$\mathbf{E}\varphi \to \varphi$$

- neither positive nor negative introspection provable
- axiom $\texttt{B}(\mathbf{E})$ provable:

$$\varphi \to \mathbf{E}\neg\mathbf{E}\neg\varphi$$

Common knowledge: basic desiderata

- truth axiom:

$$\mathbf{C}\varphi \to \varphi \text{ should be valid}$$

- positive introspection axioms:

$$\mathbf{C}\varphi \to \mathbf{E}\mathbf{C}\varphi \text{ should be valid}$$
$$\mathbf{C}\varphi \to \mathbf{C}\mathbf{C}\varphi \text{ should be valid}$$

- negative introspection axioms:

$$\neg\mathbf{C}\varphi \to \mathbf{E}\neg\mathbf{C}\varphi \text{ should be valid}$$
$$\neg\mathbf{C}\varphi \to \mathbf{C}\neg\mathbf{C}\varphi \text{ should be valid}$$

- fixed-point axiom follows:

$$\mathtt{FP} \quad \mathbf{C}\varphi \to \mathbf{E}(\varphi \wedge \mathbf{C}\varphi)$$

Common knowledge: induction principles

- two versions
  - induction axiom schema:

    $$\texttt{GFP} \quad \mathbf{C}(\varphi \to \mathbf{E}\varphi) \to (\varphi \to \mathbf{C}\varphi)$$

  - induction rule:

    $$\texttt{RGFP} \quad \text{from } \varphi \to \mathbf{E}(\varphi \wedge \psi), \text{ infer } \varphi \to \mathbf{C}\psi$$

- intuitive in temporal logics (well-founded orderings)
- doesn't 'talk' in epistemic logics

  > *"If it is the case that φ is 'self-evident', in the sense that if it is true, then everyone knows it, and, in addition, if φ is true, then everyone knows ψ, we can show by induction that if φ is true, then so is $\mathbf{E}^k(\psi \wedge \varphi)$ for all k." [vDHvdHK15]*

- aim: find a more intuitive axiom

Recap: the axiom system with induction rule

$\mathtt{S5}(\mathbf{K})$ and $\mathtt{Def}(\mathbf{E})$, plus:

$$\begin{array}{ll} \mathtt{FP} & \mathbf{C}\varphi \to \mathbf{E}(\varphi \wedge \mathbf{C}\varphi) \\ \mathtt{RGFP} & \text{from } \varphi \to \mathbf{E}(\varphi \wedge \psi)\text{, infer } \varphi \to \mathbf{C}\psi \end{array}$$

[HM92, FHMV95]

- sound and complete for S5 models
  - rule of necessitation $\mathtt{RN}(\mathbf{C})$ derivable
  - axioms $\mathtt{K}(\mathbf{C})$, $\mathtt{T}(\mathbf{C})$, $\mathtt{4}(\mathbf{C})$, $\mathtt{5}(\mathbf{C})$ provable
  - induction axiom schema $\mathtt{GFP}$ provable

Recap: the axiom system with induction axiom

$\mathtt{S5}(\mathbf{K})$ and $\mathtt{Def}(\mathbf{E})$, plus:

$$
\begin{array}{ll}
\mathtt{K}(\mathbf{C}) & \text{system K for } \mathbf{C} \\
\mathtt{FP} & \mathbf{C}\varphi \to \mathbf{E}(\varphi \wedge \mathbf{C}\varphi) \\
\mathtt{GFP} & \mathbf{C}(\varphi \to \mathbf{E}\varphi) \to (\varphi \to \mathbf{C}\varphi)
\end{array}
$$

[Leh84, HM85]

- sound and complete for S5 models
  - induction rule $\mathtt{RGFP}$ provable
  - original presentation has moreover axioms $\mathtt{T}(\mathbf{C})$, $\mathtt{4}(\mathbf{C})$, $\mathtt{5}(\mathbf{C})$
    $\Rightarrow$ redundant!

A new axiomatisation of S5 common knowledge

$\mathtt{S5}(\mathbf{K})$ and $\mathtt{Def}(\mathbf{E})$, plus:

$\mathtt{S4}(\mathbf{C})$	an axiomatisation of S4 for $\mathbf{C}$
$\mathtt{FP}_1$	$\mathbf{C}\varphi \to \mathbf{E}\varphi$
$\mathtt{GFP}_1$	$\mathbf{C}(\mathbf{E}\varphi \vee \mathbf{E}\neg\varphi) \to (\mathbf{C}\varphi \vee \mathbf{C}\neg\varphi)$

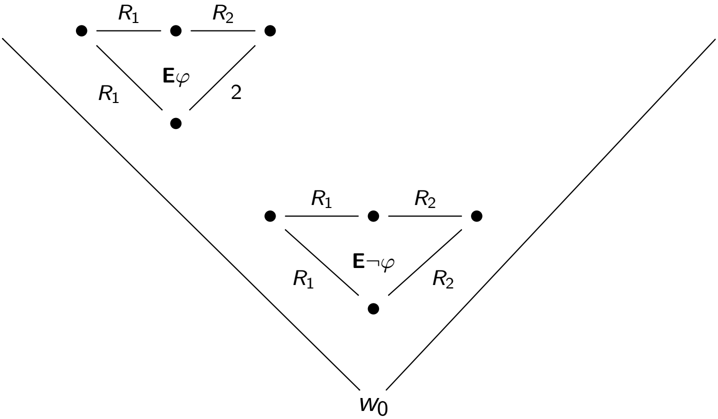
- sound for S5 models
  - $\mathtt{GFP}_1$ provable in the axiom system with induction axiom $\mathtt{GFP}$
- complete for S5 models
  - induction axiom $\mathtt{GFP}$ provable
  - proof uses $\mathtt{K}(\mathbf{C})$, $\mathtt{RN}(\mathbf{C})$,$\mathtt{T}(\mathbf{C})$ $\mathtt{4}(\mathbf{C})$
- given $\mathtt{S5}(\mathbf{K})$, $\mathtt{GFP}_1$ equivalent to the (a priori stronger):

$$\mathtt{GFP}_2 \quad \mathbf{C}\Big(\bigwedge_{i\in Agt} \mathbf{K}_i\varphi \vee \mathbf{K}_i\neg\varphi\Big) \to (\mathbf{C}\varphi \vee \mathbf{C}\neg\varphi)$$

Semantics of the new axiom

$$\boxed{\texttt{GFP}_1: \quad \mathbf{C}(\mathbf{E}\varphi \vee \mathbf{E}\neg\varphi) \to (\mathbf{C}\varphi \vee \mathbf{C}\neg\varphi)}$$

excludes:



(big cone $= R_{\mathbf{E}}^*(w_0)$; triangles $= R_{\mathbf{E}}$-accessible worlds)

Soundness: proof of $\mathtt{GFP}_1$

Proposition

$\mathtt{GFP}_1$ *is provable from* $\mathtt{GFP}$.

Proof.

- premise of $\mathtt{GFP}_1$ logically stronger than that of $\mathtt{GFP}$
  - relies on $\mathtt{T}(\mathbf{K})$ (or rather, its consequence $\mathtt{T}(\mathbf{E})$)
- case analysis:
  1. $\big(\varphi \wedge \mathbf{C}(\mathbf{E}\varphi \vee \mathbf{E}\neg\varphi)\big) \to (\varphi \wedge \mathbf{C}(\varphi \to \mathbf{E}\varphi))$ — $\mathtt{T}(\mathbf{E})$, $\mathtt{K}(\mathbf{C})$
  2. $\big(\varphi \wedge \mathbf{C}(\varphi \to \mathbf{E}\varphi)\big) \to \mathbf{C}\varphi$ — $\mathtt{GFP}$
  3. $\big(\varphi \wedge \mathbf{C}(\mathbf{E}\varphi \vee \mathbf{E}\varphi)\big) \to (\mathbf{C}\varphi \vee \mathbf{C}\neg\varphi)$ — from 1, 2
  4. $\big(\neg\varphi \wedge \mathbf{C}(\mathbf{E}\varphi \vee \mathbf{E}\neg\varphi)\big) \to (\mathbf{C}\varphi \vee \mathbf{C}\neg\varphi)$ — from 3 by uniform substitution of φ by $\neg\varphi$
  5. $\mathbf{C}(\mathbf{E}\varphi \vee \mathbf{E}\neg\varphi) \to (\mathbf{C}\varphi \vee \mathbf{C}\neg\varphi)$ — from 3, 4

Completeness: a key lemma

Lemma

The schema $\mathbf{C}(\varphi \to \mathbf{E}\varphi) \to \mathbf{C}(\neg\varphi \to \mathbf{E}\neg\varphi)$ *is provable from the axiom schemas* $\mathtt{K}(\mathbf{C})$, $\mathtt{4}(\mathbf{C})$, $\mathtt{RN}(\mathbf{C})$, $\mathtt{T}(\mathbf{C})$, *and* $\mathtt{FP}$.

Proof.

1. $\mathbf{C}(\varphi \to \mathbf{E}\varphi) \to \mathbf{E}(\varphi \to \mathbf{E}\varphi)$ — by $\mathtt{FP}$, $\mathtt{T}(\mathbf{C})$, $\mathtt{K}(\mathbf{E})$
2. $\mathbf{E}(\varphi \to \mathbf{E}\varphi) \to (\mathbf{E}\neg\mathbf{E}\varphi \to \mathbf{E}\neg\varphi)$ — by $\mathtt{K}(\mathbf{E})$
3. $\neg\varphi \to \mathbf{E}\neg\mathbf{E}\varphi$ — $\mathtt{B}(\mathbf{E})$
4. $\mathbf{C}(\varphi \to \mathbf{E}\varphi) \to (\neg\varphi \to \mathbf{E}\neg\varphi)$ — from 1, 2, 3
5. $\mathbf{CC}(\varphi \to \mathbf{E}\varphi) \to \mathbf{C}(\neg\varphi \to \mathbf{E}\neg\varphi)$ — from 4 by $\mathtt{RN}(\mathbf{C})$ and $\mathtt{K}(\mathbf{C})$
6. $\mathbf{C}(\varphi \to \mathbf{E}\varphi) \to \mathbf{CC}(\varphi \to \mathbf{E}\varphi)$ — $\mathtt{4}(\mathbf{C})$
7. $\mathbf{C}(\varphi \to \mathbf{E}\varphi) \to \mathbf{C}(\neg\varphi \to \mathbf{E}\neg\varphi)$ — from 5 and 6

Completeness: proof of GFP

Proposition

GFP *is provable from* $\mathtt{GFP}_1$.

Proof.

1. $\mathbf{C}(\mathbf{E}\varphi \vee \mathbf{E}\neg\varphi) \to (\mathbf{C}\varphi \vee \mathbf{C}\neg\varphi)$ — $\mathtt{GFP}_1$
2. $\big(\mathbf{C}(\varphi \to \mathbf{E}\varphi) \wedge \mathbf{C}(\neg\varphi \to \mathbf{E}\neg\varphi)\big) \to \mathbf{C}(\mathbf{E}\varphi \vee \mathbf{E}\neg\varphi)$ — by $\mathtt{RN}(\mathbf{C})$ and $\mathtt{K}(\mathbf{C})$
3. $\big(\mathbf{C}(\varphi \to \mathbf{E}\varphi) \wedge \mathbf{C}(\neg\varphi \to \mathbf{E}\neg\varphi)\big) \to (\mathbf{C}\varphi \vee \mathbf{C}\neg\varphi)$ — from 1 and 2
4. $\mathbf{C}(\varphi \to \mathbf{E}\varphi) \to \mathbf{C}(\neg\varphi \to \mathbf{E}\neg\varphi)$ — key lemma
5. $\mathbf{C}(\varphi \to \mathbf{E}\varphi) \to (\mathbf{C}\varphi \vee \mathbf{C}\neg\varphi)$ — from 3, 4
6. $\mathbf{C}(\varphi \to \mathbf{E}\varphi) \to (\mathbf{C}\varphi \vee \neg\varphi)$ — from 5 by $\mathtt{T}(\mathbf{C})$

Example: compatriots [LH15]

- in a conference break, two Dutch talk together in English, not knowing that they are compatriots

$$\neg(\mathbf{K}_1 d_2 \vee \mathbf{K}_1 \neg d_2) \wedge \neg(\mathbf{K}_2 d_1 \vee \mathbf{K}_2 \neg d_1) \tag{1}$$

- a third person tells them: "hey, you're compatriots"

$$\mathbf{C}(d_1 \leftrightarrow d_2) \tag{2}$$

- background knowledge:

$$\mathbf{C} \bigwedge_i \left((d_i \to \mathbf{K}_i d_i) \wedge (\neg d_i \to \mathbf{K}_i \neg d_i)\right) \tag{3}$$

- implies common knowledge that both are compatriots:

$$(2) \wedge (3) \to (\mathbf{C}(d_1 \wedge d_2) \vee \mathbf{C}(\neg d_1 \wedge \neg d_2)) \tag{4}$$

Example: compatriots (ctd.)

- common knowledge obtained through deduction, using $\mathtt{GFP}_1$
- proof:
  1. $(2) \wedge (3) \to \mathbf{C}\big(\mathbf{E}(d_1 \wedge d_2) \vee \mathbf{E}(\neg d_1 \wedge \neg d_2)\big)$
     by $\mathtt{4}(\mathbf{C})$, $\mathtt{FP}_1$, $\mathtt{K}(\mathbf{C})$
  2. $\mathbf{C}\big(\mathbf{E}(d_1 \wedge d_2) \vee \mathbf{E}(\neg d_1 \wedge \neg d_2)\big) \to \big(\mathbf{C}(d_1 \wedge d_2) \vee \mathbf{C}(\neg d_1 \wedge \neg d_2)\big)$
     consequence of $\mathtt{GFP}_1$
  3. $(2) \wedge (3) \to \big(\mathbf{C}(d_1 \wedge d_2) \vee \mathbf{C}(\neg d_1 \wedge \neg d_2)\big)$
     from 1, 2

- induction principles not so easy to apply
  - group version of the omniscience problem?
  - implicit vs. explicit common knowledge [LH15]
- often common knowledge cannot be deduced
  - cf. consecutive numbers puzzle

Commonly knowing whether

- definable from 'knowing that' operators:
  - $\mathbf{Kif}_i\varphi = \mathbf{K}_i\varphi \vee \mathbf{K}_i\neg\varphi$ "i knows whether φ"
  - $\mathbf{Eif}\varphi = \mathbf{E}\varphi \vee \mathbf{E}\neg\varphi$ "it is shared knowledge whether φ"
  - $\mathbf{Cif}\varphi = \mathbf{C}\varphi \vee \mathbf{C}\neg\varphi$ "it is common knowledge whether φ"
- the other way round:
  - $\mathbf{C}\varphi = \varphi \wedge \mathbf{Cif}\varphi$
  - $\mathbf{E}\varphi = \ldots$
  - $\mathbf{C}\varphi = \ldots$
- easy axiomatisation
  1. axiomatisation of $\mathbf{Kif}_i$ [FWvD15]
  2. $\mathbf{Eif}\varphi \leftrightarrow \bigwedge_{i \in Agt} \mathbf{Kif}_i\varphi$ (N.B.: does not hold for belief!)
  3. standard axiomatisation of $\mathbf{C}$, substituting $\mathbf{C}\varphi$ by $\varphi \wedge \mathbf{Cif}\varphi$
     - induction axiom GFP becomes:

$$\big((\varphi \to \mathbf{Eif}\varphi) \wedge \mathbf{Cif}(\varphi \to \mathbf{Eif}\varphi)\big) \to (\varphi \to \mathbf{Cif}\varphi)$$

$\Rightarrow$ not easy to parse

A nicer axiomatisation of commonly knowing whether

- $\mathtt{FP}_1$ and $\mathtt{GFP}_1$ become:

$$\begin{array}{ll} \mathtt{FP}_2 & \textbf{Cif}\varphi \to \textbf{Eif}\varphi \\ \mathtt{GFP}_2 & \textbf{Cif}\,\textbf{Eif}\varphi \to \big(\textbf{Eif}\varphi \to \textbf{Cif}\varphi\big) \end{array}$$

Conclusion

- $\mathtt{GFP}_1$ 'talks to us'
  - more intuitive than the standard induction principles
- what about logics weaker than S5?
  - $\mathtt{GFP}_2$ surely not a reasonable principle of common belief:
    - common belief that each of us has an opinion about φ does not imply common belief about φ
  - the weaker $\mathtt{GFP}_1$ not reasonable either!
  - counterexample: misunderstanding in conversation
    - $\mathbf{B}_1\,\mathbf{CB}\,p \wedge \mathbf{B}_2\,\mathbf{CB}\,\neg p$
    - consequence: $\mathbf{B}_1\,\mathbf{CB}\,\mathbf{EB}\,p \wedge \mathbf{B}_2\,\mathbf{CB}\,\mathbf{EB}\,\neg p$
    - consequence: $\mathbf{B}_1\,\mathbf{CB}\,(\mathbf{EB}\,p \vee \mathbf{EB}\,\neg p) \wedge \mathbf{B}_2\,\mathbf{CB}\,(\mathbf{EB}\,p \vee \mathbf{EB}\,\neg p)$
    - consequence: $\mathbf{CB}\,(\mathbf{EB}\,p \vee \mathbf{EB}\,\neg p)$

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

Jie Fan, Yanjing Wang, and Hans van Ditmarsch.
Contingency and knowing whether.
Rew. Symb. Logic, 8(1):75–107, 2015.

Joseph Y. Halpern and Yoram Moses.
A guide to the modal logics of knowledge and belief: Preliminary draft.
In *Proceedings IJCAI'85*, pages 480–490. Morgan Kaufmann, 1985.

Joseph Y. Halpern and Yoram Moses.
A guide to completeness and complexity for modal logics of knowledge and belief.
Artificial Intelligence, 54(3):319–379, 1992.

Daniel J. Lehmann.
Knowledge, common knowledge and related puzzles (extended summary).
In *Proc. PODC*, pages 62–67, 1984.

H. van Ditmarsch, J.Y. Halpern, W. van der Hoek, and B.P. Kooi.
Handbook of Epistemic Logic.
College Publications, 2015.