499 Modal and Temporal Logic

Epistemic Logic and ‘Common Knowledge’

Marek Sergot
Department of Computing
Imperial College, London

Autumn 2007

Further reading:

- R. Fagin, J. Y. Halpern, Y. Moses, and M. Y. Vardi. *Reasoning about Knowledge.* MIT Press, Cambridge, 1995.
- J-J. Ch. Meyer and W. van der Hoek. *Epistemic Logic for Artificial Intelligence and Computer Science.* Cambridge University Press, 1995.

The logic S5^n, and variations

The logic S5 $= KT5 = KT45$ is often taken as the standard logic of rational knowledge for a single agent (and $KD45$ (‘weak S5’) as the standard logic of belief).

Given a set of agents $\{1, \ldots, n\}$
The logic S5^n is the smallest modal logic in which each K_i is of type S5 (i.e., of type $KT5 = KT45 = KT4B$), i.e. the smallest modal logic containing (for each $i \in 1..n$):

RN.	$\dfrac{A}{\mathsf{K}_i A}$	
K.	$\mathsf{K}_i(A \to B) \to (\mathsf{K}_i A \to \mathsf{K}_i B)$	
T.	$\mathsf{K}_i A \to A$	*‘veridicality’* or *‘truth’*
4.	$\mathsf{K}_i A \to \mathsf{K}_i \mathsf{K}_i A$	*‘positive introspection’*
5.	$\neg \mathsf{K}_i A \to \mathsf{K}_i \neg \mathsf{K}_i A$	*‘negative introspection’*

There are just 6 modalities in S5 (see ‘reduction laws’, exercise sheet 1, e.g., $\vdash_{S5} \mathsf{K}_i{}^k A \leftrightarrow \mathsf{K}_i A$, etc):

$$\begin{array}{ccccc} \mathsf{K}_i A & \to & A & \to & \neg \mathsf{K}_i \neg A \\ \mathsf{K}_i \neg A & \to & \neg A & \to & \neg \mathsf{K}_i A \end{array}$$

A standard logic of belief: $KD45^n$ (sometimes ‘weak S5^n’)

RN.	$\dfrac{A}{\mathsf{B}_i A}$	
K.	$\mathsf{B}_i(A \to B) \to (\mathsf{B}_i A \to \mathsf{B}_i B)$	
P.	$\neg \mathsf{B}_i \bot$	*‘consistency’*
D.	$\mathsf{B}_i A \to \neg \mathsf{B}_i \neg A$	
4.	$\mathsf{B}_i A \to \mathsf{B}_i \mathsf{B}_i A$	*‘positive introspection’*
5.	$\neg \mathsf{B}_i A \to \mathsf{B}_i \neg \mathsf{B}_i A$	*‘negative introspection’*

Note that in every *normal* system the schemas P and D are inter-derivable: a normal system contains P iff it contains D.

These are *very strong* properties, whether we read them as referring to knowledge or belief. For example, it is generally accepted that negative introspection is a more demanding condition than positive introspection. Therefore many researchers argue that it is more reasonable to adopt S4^n $= KT4^n$, rather than S5^n, as the logic of knowledge (and $KD4^n$ as the logic of belief).

Logical omniscience

Notice that these are all *normal* logics and so they have, among other things:

$$\frac{A \to B}{\mathsf{K}_i A \to \mathsf{K}_i B}$$

An agent knows all the logical consequences of what it knows — one manifestation of *logical omniscience*.
Clearly this is not a property of real agents, and what they actually know.

But if modal epistemic logics do not describe what agents *actually* know, what *do* they describe?
Several possible suggested readings for $\mathsf{K}_i A$:

- “agent i knows A implicitly”
- “A follows from i’s knowledge”
- “agent i carries the information A”
- “A is agent i’s possible knowledge”

These and other possible suggestions refer to what is implicitly represented in an agent’s information state, i.e., what logically follows from its actual knowledge. They do not refer to any notion of how an agent computes knowledge or answers questions based on its knowledge. What an agent actually knows is called its explicit knowledge. We won’t be looking at possible formalisations of explicit knowledge and actual reasoning mechanisms.

Models

Given a set of agents $\{1, \ldots, n\}$

$$\mathcal{M} = \langle W, R_1, \ldots, R_n, h \rangle$$

For $S5^n$ every accessibility relation R_i is an equivalence relation.
Note that every transitive relation is symmetric if and only if it is euclidean.

Theorem

1. K^n is determined by the class of all models with n accessibility relations.
2. $T^n = KT^n$ is determined by the class of models where the n accessibility relations are all reflexive.
3. $S4^n = KT4^n$ is determined by the class of models where the n accessibility relations are all reflexive and transitive.
4. $S5^n = KT5^n = KT45^n$ is determined by the class of models where the n accessibility relations are all equivalence relations.
5. KD^n is determined by the class of models where the n accessibility relations are all serial.
6. $KD4^n$ is determined by the class of models where the n accessibility relations are all serial and transitive.
7. $KD45^n$ is determined by the class of models where the n accessibility relations are all serial, transitive, and euclidean.

Mutual knowledge – 'everyone knows'

The auxiliary operator E (to be interpreted as "everyone knows") is defined as:

$$\mathsf{E}\, A =_{\text{def}} \mathsf{K}_1\, A \wedge \ldots \wedge \mathsf{K}_n\, A$$

Or more generally where G is any non-empty subset of $\{1, \ldots, n\}$:

$$\mathsf{E}_G\, A =_{\text{def}} \bigwedge_{i \in G} \mathsf{K}_i\, A$$

$\mathsf{E}_G\, A$ – "everyone in group G knows A".
Sometimes called *mutual knowledge.*
Similarly we can define *mutual belief* – "everyone (in group G) believes A."

Given a model $\mathcal{M} = \langle W, R_1, \ldots, R_n, h \rangle$, we can define the truth conditions for $\mathsf{E}_G\, A$ as follows:

$$\begin{aligned} \mathcal{M}, w \models \mathsf{E}_G\, A \quad & \text{iff for every } i \text{ in } G,\ \mathcal{M}, w \models \mathsf{K}_i\, A \\ & \text{iff for every } i \text{ in } G, \text{ for every } w',\ w\, R_i\, w' \text{ implies } \mathcal{M}, w' \models A \\ & \text{iff for every } w',\ w\, R_{\mathsf{E}_G}\, w' \text{ implies } \mathcal{M}, w' \models A \end{aligned}$$

where $R_{\mathsf{E}_G} = R_{i_1} \cup \cdots \cup R_{i_m}$ for $G = \{i_1, \ldots, i_m\}$. $R_{\mathsf{E}_G} = \bigcup_{i \in G} R_i$.

E_G is evaluated on a relation (the relation R_{E_G}). It follows that each E_G is normal:

$$\text{RN.} \qquad \frac{A}{\mathsf{E}_G\, A}$$

$$\text{K.} \qquad \mathsf{E}_G(A \rightarrow B) \rightarrow (\mathsf{E}_G\, A \rightarrow \mathsf{E}_G\, B)$$

Further, if each R_i is reflexive, then clearly $R_1 \cup \cdots \cup R_n$ is also reflexive (it is enough that one of the R_i is reflexive), and so we also have:

$$\text{T.} \quad \mathsf{E}_G\, A \rightarrow A$$

However, it is easy to check:

- $R_1 \cup \cdots \cup R_n$ is not necessarily transitive even if all the R_i are transitive;
- $R_1 \cup \cdots \cup R_n$ is not necessarily euclidean even if all the R_i are euclidean.

And so the following are *not* valid:

$$\not\models \mathsf{E}_G\, A \rightarrow \mathsf{E}_G\, \mathsf{E}_G\, A$$
$$\not\models \neg\, \mathsf{E}_G\, A \rightarrow \mathsf{E}_G\, \neg\, \mathsf{E}_G\, A$$

However, the union of a set of symmetric relations is also symmetric, and so if all the R_i are symmetric (as they are if all are equivalence relations) then the following schema is valid:

$$\text{B.} \quad A \rightarrow \mathsf{E}_G\, \neg\, \mathsf{E}_G\, \neg A$$

It is also easy to see that

$$\models \mathsf{E}_G\, A \rightarrow \mathsf{E}_{G'}\, A \qquad \text{when } G' \subseteq G$$

One can construct a representation of mutual belief ("everyone in group G believes") in similar fashion.
Notice that the union of a set of serial relations is also serial. And so, e.g.:

$$\vdash_{KD45^n} \neg\, \mathsf{E}_G\, \bot$$
$$\vdash_{KD45^n} \mathsf{E}_G\, A \rightarrow \neg\, \mathsf{E}_G\, \neg A$$

Distributed knowledge

(Not so interesting in my opinion.)
If mutual knowledge of a group of agents corresponds to the union of the accessibility relations $R_1 \cup \cdots \cup R_n$, what kind of knowledge corresponds to the intersection $R_1 \cap \cdots \cap R_n$?

Given a model $\mathcal{M} = \langle W, R_1, \ldots, R_n, h \rangle$, define the truth conditions for $\mathsf{D}_G\, A$ as follows:

$$\begin{array}{ll} \mathcal{M}, w \models \mathsf{D}_G\, A & \text{iff for every } w' \text{ such that } w\, R_i\, w' \text{ for every } i \in G \text{ we have } \mathcal{M}, w' \models A \\ & \text{iff for every } w',\ w\, R_{\mathsf{D}_G}\, w' \text{ implies } \mathcal{M}, w' \models A \end{array}$$

where $R_{\mathsf{D}_G} = R_{i_1} \cap \cdots \cap R_{i_m}$ for $G = \{i_1, \ldots, i_m\}$. $R_{\mathsf{D}_G} = \bigcap_{i \in G} R_i$.

Easy to see that the following schema is valid:

$$\models \mathsf{K}_i\, A \to \mathsf{D}_G\, A \qquad \text{for every } i \in G$$

Or in other words: $\models \bigvee_{i \in G} \mathsf{K}_i\, A \to \mathsf{D}_G\, A$

This is easy to check because $\bigcap_{j \in G} R_j \subseteq R_i$ for every $i \in G$.

But the following is *not* valid:

$$\not\models \mathsf{D}_G\, A \to \bigvee_{i \in G} \mathsf{K}_i\, A$$

So $\mathsf{D}_G\, A$ means that group G 'knows' A if they could somehow *pool* their information — even when no i in G *individually* knows A.

Clearly: $\models \mathsf{D}_G\, A \to \mathsf{D}_{G'}\, A$ when $G \subseteq G'$

D_G is interpreted on a relation (the relation $R_{\mathsf{D}_G} = \bigcap_{i \in G} R_i$).
It follows that every D_G is normal. The logic of distributed knowledge has:

$$\text{RN.} \qquad \frac{A}{\mathsf{D}_G\, A}$$

$$\text{K.} \qquad \mathsf{D}_G(A \to B) \to (\mathsf{D}_G\, A \to \mathsf{D}_G\, B)$$

Also:

- if each R_i is reflexive then $\bigcap_{i \in G} R_i$ is reflexive;
- if each R_i is symmetric then $\bigcap_{i \in G} R_i$ is symmetric;
- if each R_i is transitive then $\bigcap_{i \in G} R_i$ is transitive.

And so e.g. the logic S5^n with distributed knowledge has:

$$\begin{array}{ll} \text{T.} & \mathsf{D}_G\, A \to A \\ \text{4.} & \mathsf{D}_G\, A \to \mathsf{D}_G\, \mathsf{D}_G\, A \\ \text{5.} & \neg\, \mathsf{D}_G\, A \to \mathsf{D}_G\, \neg\, \mathsf{D}_G\, A \end{array}$$

Distributed knowledge is not very interesting, in my opinion. However, there is a recently established (August 2007!) connection to the logic of *collective action*.
Roughly: read $\mathsf{D}_i A$ as 'A is a necessary consequence of what i does'.
Then $\mathsf{D}_G\, A$ represents a kind of *collective* action by the group G: A is a necessary consequence of the group G's collective actions, though not a necessary consequence of what any of the individual members in G does.'

Some useful observations

Consider models/frames

$$\mathcal{M} = \langle W, R_1, R_2, \ldots \rangle$$

with $\Box_1$ and $\Box_2$ interpreted on R_1 and R_2 respectively.

If $R_1 \subseteq R_2$ then

- $\models \Diamond_1 A \to \Diamond_2 A$
- $\models \Box_2 A \to \Box_1 A$

It is easy to check that $\Box_2 A \to \Box_1 A$ is canonical for $R_1 \subseteq R_2$.

Mutual knowledge ('everyone knows')

$$R_{\mathsf{E}_G} = R_1 \cup \cdots \cup R_n$$

- $R_{\mathsf{E}_G} \subseteq R_1 \cup \cdots \cup R_n$ gives $\models \mathsf{K}_1\, A \wedge \cdots \wedge \mathsf{K}_n\, A \to \mathsf{E}_G\, A$.
- $R_1 \cup \cdots \cup R_n \subseteq R_{\mathsf{E}_G}$ gives $\models \mathsf{E}_G \to \mathsf{K}_1\, A \wedge \cdots \wedge \mathsf{K}_n\, A$.

$\mathsf{E}_G \leftrightarrow \mathsf{K}_1\, A \wedge \cdots \wedge \mathsf{K}_n\, A$ is canonical for $R_{\mathsf{E}_G} = R_1 \cup \cdots \cup R_n$.

Distributed knowledge

$$R_{\mathsf{D}_G} = R_1 \cap \cdots \cap R_n$$

- $R_1 \cap \cdots \cap R_n \subseteq R_i$ so $\models \mathsf{K}_i\, A \to \mathsf{D}_G\, A$.

But $\models \mathsf{K}_i\, A \to \mathsf{D}_G\, A$ just implies

- $R_i \subseteq R_{\mathsf{D}_G}$ for every $i \in G$
- $R_1 \cap \cdots \cap R_n \subseteq R_{\mathsf{D}_G}$

$\mathsf{K}_1\, A \vee \cdots \vee \mathsf{K}_n\, A \to \mathsf{D}_G\, A$ is canonical for $R_1 \cap \cdots \cap R_n \subseteq R_{\mathsf{D}_G}$.

Finally (useful in a minute)

$$\mathcal{M} = \langle W, R_1, R_2, \ldots \rangle$$

$$\mathcal{M}, w \models \Box_1 \Box_2 A \quad \text{iff } \mathcal{M}, w' \models A \text{ for all } (w, w') \in R_1 \circ R_2$$

where $R_1 \circ R_2$ is the *composition* of relations R_1 and R_2:

$$(w, w') \in R_1 \circ R_2 \quad \text{iff there exists } w'' \text{ such that } (w, w'') \in R_1 \text{ and } (w'', w') \in R_2$$

Common knowledge

Basic idea: it is *common knowledge* in group G that A when everyone in group G knows A, and everyone in group G knows everyone in group G knows A, and everyone in group G knows everyone in group G knows everyone in group G knows A, etc, etc.

$$\mathsf{C}_G\, A \;\leftrightarrow\; \mathsf{E}_G\, A \;\wedge\; \mathsf{E}_G\,\mathsf{E}_G\, A \;\wedge \cdots \wedge\; {\mathsf{E}_G}^k\, A \;\wedge \ldots$$

But the above is an infinitely long conjunction, and hence is not a well formed formula.

Given a model $\mathcal{M} = \langle W, R_1, \ldots, R_n, h \rangle$, we can define the truth conditions for $\mathsf{C}_G\, A$ as follows:

$$\mathcal{M}, w \models \mathsf{C}_G\, A \quad \text{iff } \mathcal{M}, w \models {\mathsf{E}_G}^k\, A \text{ for every } k \geq 1$$

Reminder

When R and S are both binary relations on a set W their *composition* $R \circ S$ is defined as follows:

$$(w, w') \in R \circ S \quad \text{iff there exists } w'' \text{ such that } (w, w'') \in R \text{ and } (w'', w') \in S$$

Usual notation: $R^1 = R, \quad R^2 = R \circ R, \quad \ldots, \quad R^{k+1} = R \circ R^k = R^k \circ R, \quad \ldots.$

So R^k can be seen as the set of all paths of length k of R (or rather, the set of pairs of elements of W that are connected by paths of length k of R).

The *transitive closure* R^+ of a binary relation R is the smallest (set inclusion) transitive relation that contains R. And ...

$$R^+ \quad = \quad R^1 \cup R^2 \cup \cdots \cup R^k \cup \ldots \quad = \quad \bigcup_{k \geq 1} R^k$$

So, given a model $\mathcal{M} = \langle W, R_1, \ldots, R_n, h \rangle$, we can also define the truth conditions for $\mathsf{C}_G\, A$ as follows:

$$\begin{aligned} \mathcal{M}, w \models \mathsf{C}_G\, A \quad & \text{iff } \mathcal{M}, w \models {\mathsf{E}_G}^k\, A \text{ for every } k \geq 1 \\ & \text{iff for every } w',\ w\, R^k_{\mathsf{E}_G}\, w' \text{ implies } \mathcal{M}, w' \models A, \text{ for every } k \geq 1 \\ & \text{iff for every } w',\ w\, R_{\mathsf{C}_G}\, w' \text{ implies } \mathcal{M}, w' \models A \end{aligned}$$

where $R_{\mathsf{C}_G} = R^+_{\mathsf{E}_G}$, the *transitive closure* of R_{E_G}.

For $G = \{i_1, \ldots, i_m\}$

$$\begin{aligned} R_{\mathsf{E}_G} &= (R_{i_1} \cup \ldots R_{i_m}) \\ R_{\mathsf{C}_G} &= (R_{i_1} \cup \ldots R_{i_m})^+ = \bigcup_{k \geq 1} (R_{i_1} \cup \ldots R_{i_m})^k \end{aligned}$$

C_G is interpreted on a relation (the relation $R_{\mathsf{C}_G} = R^+_{\mathsf{E}_G}$).
It follows that every C_G is normal. Also:

- the transitive closure of a reflexive relation is also reflexive; if the R_i are reflexive, R_{E_G} is reflexive and so is R_{C_G};
- the transitive closure of a symmetric relation is also symmetric; if the R_i are transitive and euclidean they are symmetric; R_{E_G} is symmetric and so is R_{C_G};
- the transitive closure of a relation is obviously transitive; a transitive relation that is symmetric is also euclidean; so if the R_i are transitive and euclidean, R_{C_G} is symmetric and therefore also euclidean.

And so e.g. the logic $\mathrm{S5}^n_{\mathsf{C}}$ has:

RN.	$\dfrac{A}{\mathsf{C}_G\, A}$
K.	$\mathsf{C}_G(A \to B) \to (\mathsf{C}_G\, A \to \mathsf{C}_G\, B)$
T.	$\mathsf{C}_G\, A \to A$
4.	$\mathsf{C}_G\, A \to \mathsf{C}_G\, \mathsf{C}_G\, A$
5.	$\neg\, \mathsf{C}_G\, A \to \mathsf{C}_G\, \neg\, \mathsf{C}_G\, A$

It is also easy to see that

$$\vdash_{\mathrm{S5}^n_{\mathsf{C}}} \mathsf{C}_G\, A \to \mathsf{C}_{G'}\, A \qquad \text{when } G' \subseteq G$$

And obviously

- $\vdash_{\mathrm{S5}^n_{\mathsf{C}}} \mathsf{C}_G\, A \to \mathsf{E}_G\, A$
- $\vdash_{\mathrm{S5}^n_{\mathsf{C}}} \mathsf{C}_G\, A \to \mathsf{C}_G\, \mathsf{E}_G\, A$
- $\vdash_{\mathrm{S5}^n_{\mathsf{C}}} \mathsf{C}_G\, A \to \mathsf{C}_G\, \mathsf{K}_i\, A$

etc, etc.

One can construct a representation of *common belief* ('it is a common belief in group G that") in similar fashion.
The transitive closure of a serial relation is also serial. And so:

$$\begin{aligned} & \vdash_{KD45^n_{\mathsf{C}}} \neg\, \mathsf{C}_G \perp \\ & \vdash_{KD45^n_{\mathsf{C}}} \mathsf{C}_G\, A \to \neg\, \mathsf{C}_G\, \neg A \end{aligned}$$

Axioms

Logics of common knowledge can be axiomatized on the basis of the corresponding epistemic logics by adding suitable axiom schemata and inference rules. The following axiomatization is due to Halpern and Moses.

FP. $\mathsf{C}_G A \to \mathsf{E}_G(A \land \mathsf{C}_G A)$ *'Fixpoint axiom'*

RI. $\dfrac{A \to \mathsf{E}_G(A \land B)}{A \to \mathsf{C}_G B}$ *'Rule of Induction'*

Various other axiomatizations exist. (There is no need to memorise the above.)
The rule RI is equivalent to the following (which is perhaps clearer):

$$\mathrm{RI}' \quad \frac{A \to \mathsf{E}_G(A \land B)}{A \to \mathsf{C}_G(A \land B)}$$

It is easy to show the above are sound (with respect to the class of models in which $R_{\mathsf{C}_G} = R^+_{\mathsf{E}_G}$.)
For the schema FP, notice that

- $\mathsf{C}_G A \to \mathsf{E}_G A$ is validated by $R_{\mathsf{E}_G} \subseteq R^+_{\mathsf{E}_G}$;
- $\mathsf{C}_G A \to \mathsf{E}_G \mathsf{C}_G A$ is validated by $R_{\mathsf{E}_G} \circ R^+_{\mathsf{E}_G} \subseteq R^+_{\mathsf{E}_G}$.

(Check it. Easy.)

For RI′, suppose $\mathcal{M} \models A \to \mathsf{E}_G(A \land B)$. Suppose $\mathcal{M}, w \models A$. Now show $\mathcal{M}, w \models \mathsf{C}_G(A \land B)$: show by induction on k that $\mathcal{M}, w' \models A \land B$ for every $(w, w') \in R^k_{\mathsf{E}_G}$, for every $k \geq 1$.

One can also show completeness by the canonical model method. (There are a couple of little fiddly details, which I omit. See e.g. the book by Fagin, Halpern,Moses, and Vardi.)

(There is no need to *memorise* the axiomatisation. It is included for your interest.)