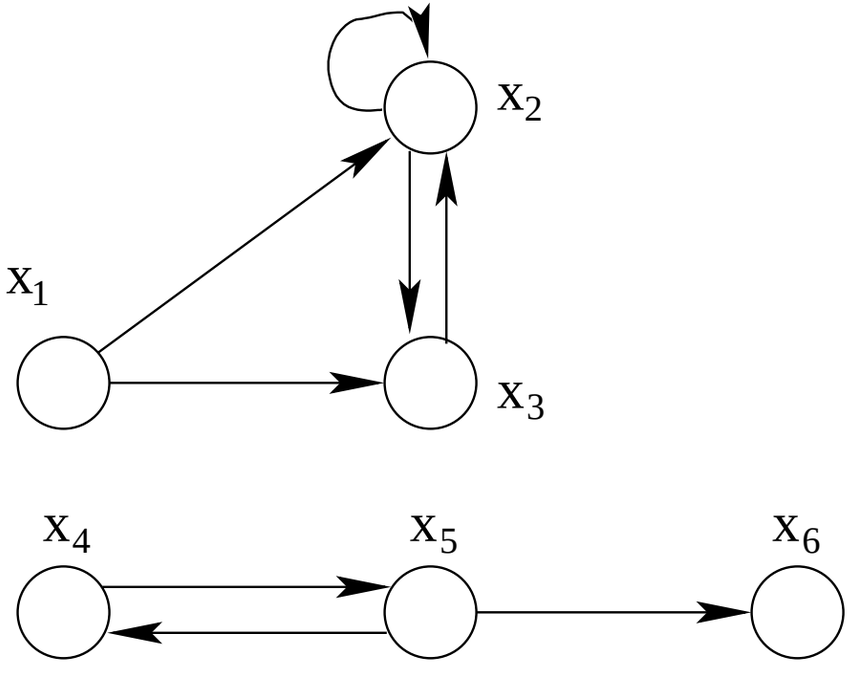
Modal Logic (2)

- There appeared to be a correspondence between the validity of $\Box\Phi \rightarrow \Phi$ and the property that the accessibility relation R is reflexive. The connection between them is that both relied on the intuition that anything which is known by an agent is true.
- There also seemed to be a correspondence between $\Box\Phi \rightarrow \Box\Box\Phi$ and R being transitive; they both seem to assert the property of ***positive introspection***, i.e. that which is known is known to be known.
- To every formula scheme there corresponds a property of R.
- This relationship helps one to understand the logic being studied.
  - If you believe that a certain formula scheme should be accepted in the system of modal logic you are engineering, then it is worth looking at the corresponding property of R and checking that that makes sense for the application.
  - For some formulas it may seem difficult to understand what they mean, so looking at the corresponding property of R can help.

Frames

Definition: A ***frame*** $\mathcal{F} = (W, R)$ is a set W (of worlds) and a binary relation R on W.

- A frame is like a Kripke model, except that it has no labelling function.
- From any model we can extract a frame, by just forgetting about the labelling function; the figure below shows a Kripke model and its frame.
- A frame is just a set of "points" and the relationship between them. It has no information about what atomic formulas are true at the various points.
- It is useful to say sometimes that the frame, as a whole, satisfies a formula.



Formula Satisfied by a Frame

Definition: A frame $\mathcal{F} = (W, R)$ satisfies a formula of basic modal logic Φ, written $\mathcal{F} \models \Phi$, if for each labeling function L and each $w \in W$, we have $\mathcal{M}, w \Vdash \Phi$, where $\mathcal{M} = (W, R, L)$.

- If a frame satisfies a formula, then it also satisfies every substitution instance of that formula.
- For example, the model of the figure given below satisfies $p \vee \Diamond p \vee \Diamond\Diamond p$, but it does not satisfy every instance of $\Phi \vee \Diamond\Phi \vee \Diamond\Diamond\Phi$;
- On the other hand, for the corresponding frame, x_6 does not satisfy $q \vee \Diamond q \vee \Diamond\Diamond q$; but because frames do not contain any information about the truth or falsity of propositional atoms, they can't distinguish between different atoms; so, if a frame satisfies a formula, it also satisfies the formula scheme obtained by substituting the p, q etc. by Φ, Ψ etc.

Example

The frame in the figure satisfies $\Box p \rightarrow p$. To see this, we have to consider any labeling function of the frame and show that each world satisfies the formula for each labeling.

Generic argument: Let x be any world. Suppose $x \Vdash \Box p$; we want to show $x \Vdash p$. We know that $R(x,x)$ because each x is related to itself. It follows that $x \Vdash p$.

- Our frame satisfies any formula of this shape, i.e. it satisfies the formula scheme $\Box\Phi \rightarrow \Phi$.
- The frame does not satisfy the formula $\Box p \rightarrow \Box\Box p$. Take the labeling $L(x_4) = \{p\}, L(x_5) = \{p\}, L(x_6) = \emptyset$; then $x_4 \Vdash \Box p$, but $x_4 \nVdash \Box\Box p$.

First Correspondence Theorem

Theorem: Let $\mathcal{F} = (W, R)$ be a frame.

1. The following statements are equivalent:

   - R is reflexive;
   - F satisfies $\Box\Phi \rightarrow \Phi$;
   - F satisfies $\Box p \rightarrow p$;

2 The following statements are equivalent:

   - R is transitive;
   - $\mathcal{F}$ satisfies $\Box\Phi \rightarrow \Box\Box\Phi$;
   - $\mathcal{F}$ satisfies $\Box p \rightarrow \Box\Box p$.

Proof: For each of the cases, there are three things to prove: (a) that, if R has the property, then the frame satisfies the formula scheme; and (b) that, if the frame satisfies the formula scheme then it satisfies the instance of it; and (c) that, if the frame satisfies the formula, then R has the property.

Proof of the First Correspondence Theorem (1)

(a) Suppose R is reflexive. Let L be a labelling function, so now $\mathcal{M} = (W, R, L)$ is a model of basic modal logic. We need to show $\mathcal{M} \models \Box\Phi \rightarrow \Phi$. That means we need to show $x \Vdash \Box\phi \rightarrow \Phi$ for any $x \in W$, so pick any x. Suppose $x \Vdash \Box\Phi$; since $R(x,x)$, it immediately follows from the clause for $\Box$ that $x \Vdash p$. Therefore, we have shown $x \Vdash \Box\Phi \rightarrow \Phi$.

(b) We just set Φ to be p.

(c) Suppose the frame satisfies $\Box p \rightarrow p$. Take any x. We show $R(x,x)$. Take a labelling function L such that $p \notin L(x)$ and $p \in L(y)$ for all worlds y except x. Suppose we don't have $R(x,x)$. Then, $x \Vdash \Box p$, since all the worlds related to x satisfy p but since $\mathcal{F}$ satisfies $\Box p \rightarrow p$, it follows that $x \Vdash \Box p \rightarrow p$; therefore, putting $x \Vdash \Box p$ and $x \Vdash \Box p \rightarrow p$ together, we get $x \Vdash p$. This is a contradiction, since we said that $p \notin L(x)$. We got this contradiction just by assuming that we didn't have $R(x,x)$. So this assumption cannot be made. It must be that $R(x,x)$ holds in our frame!

Proof of the First Correspondence Theorem (2)

(a) Suppose R is transitive. Let L be a labelling function and $\mathcal{M} = (W, R, L)$. We need to show $M \Vdash \Box\Phi \rightarrow \Box\Box\Phi$. That means we need to show $x \Vdash \Box\Phi \rightarrow \Box\Box\Phi$, for any $x \in W$, so pick any x. Suppose $x \Vdash \Box\Phi$; we need to show $x \Vdash \Box\Box\Phi$. That is, using the clause for $\Box$, any y such that $R(x,y)$ satisfies $\Box\Phi$; that is, for any y, z with $R(x,y)$ and $R(y,z)$, we have $z \Vdash \Phi$. Suppose we did have y and z with $R(x,y)$ and $R(y,z)$. By the fact that R is transitive, we obtain $R(x,z)$. But we're supposing that $x \Vdash \Box\Phi$, so (clause for $\Box$) that means $z \Vdash \Phi$, which is what we needed to prove.

(b) Again, just set Φ to be p.

(c) Suppose the frame satisfies $\Box p \rightarrow \Box\Box p$. Take any x, y and z with $R(x,y)$ and $R(y,z)$; we are going to show $R(x,z)$. Take a labelling function L such that $p \notin L(z)$ and $p \in L(w)$ for all worlds w except z. Suppose we don't have $R(x,z)$; then $x \Vdash \Box p$, since $w \Vdash p$ for all $w \neq z$. Using the axiom $\Box p \rightarrow \Box\Box p$, it follows that $x \Vdash \Box\Box p$; i.e., $y \Vdash \Box p$ (since $R(x,y)$), i.e., $z \Vdash p$ (since $R(y,z)$). So we get a contradiction. Thus, we must have $R(x,z)$.

Second Correspondence Theorem

Theorem: A frame $\mathcal{F} = (W, R)$ satisfies a formula scheme in the table below iff R has the corresponding property in that table.

The names of the formulas in the left-hand column are historical, but have stuck and are still used widely in the literature.

name	formula scheme	property of R
T	$\Box\phi \rightarrow \phi$	reflexive
B	$\phi \rightarrow \Box\Diamond\phi$	symmetric
D	$\Box\phi \rightarrow \Diamond\phi$	serial
4	$\Box\phi \rightarrow \Box\Box\phi$	transitive
5	$\Diamond\phi \rightarrow \Box\Diamond\phi$	Euclidean
	$\Box\phi \leftrightarrow \Diamond\phi$	functional
	$\Box(\phi \wedge \Box\phi \rightarrow \psi) \vee \Box(\psi \wedge \Box\psi \rightarrow \phi)$	linear

Some Modal Logics

We build a modal logic by picking and choosing among formula schemes, according to the application at hand.

Definition: A modal logic L is a subset of formulas of basic modal logic, with the following properties:

1. L is closed under propositional logic. That is, anything which can be derived from members of L using propositional logic is itself a member of L.
2. L contains all instances of the formula scheme K:

$$\Box(\Phi \rightarrow \Psi) \rightarrow (\Box\Phi \rightarrow \Box\Psi).$$

3. L is closed under the ***rule of necessitation***; this says that, if $\Phi \in L$, then also $\Box\Phi \in L$.
4. L is closed under taking substitution instances; meaning that, if Φ is in L, then any substitution instance of Φ is also in L.

To build a modal logic, choose the formula schemes which you would like to have inside it. These are called the axioms of the logic. Then, 'close' it under the conditions of the definition.

The Modal Logic K

- The weakest modal logic doesn't have any 'optional' formula schemes.
- It just contains propositional logic and all instances of the formula scheme K, together with other formulas which come from applying conditions 3 and 4 of the definition on slide 9.
- The name K is given to this logic (as well as being given to the formula scheme K).

The Modal Logic $KT45$

- Modal logic $KT45$ (also called $S5$ in the literature), adds three extra axioms.
- This is used to reason about knowledge; $\Box\Phi$ means that the agent Q knows Φ. The axioms T, 4 and 5, respectively, tell us that

  T. ***Truth:*** the agent Q knows only true things.

  4. ***Positive introspection:*** if the agent Q knows something, then she knows that she knows it.

  5. ***Negative introspection:*** if the agent Q doesn't know something, then she knows that she doesn't know it.

- The formula scheme K means ***logical omniscience:*** the agent's knowledge is closed under logical consequence.
- Note that these properties represent idealisations of knowledge.
- Human knowledge has none of these properties! Even computer agents may not have them all.

$KT45$ Properties

- The semantics of the logic $KT45$ must consider only relations R which are: ***reflexive*** (T), ***transitive*** (4), and ***Euclidean*** (5).
- A relation is reflexive, transitive and Euclidean iff it is reflexive, transitive and symmetric, i.e. if it is an equivalence relation.
- $KT45$ is simpler than K in the sense that it has few essentially different ways of composing modalities.

Theorem Any sequence of modal operators and negations in $KT45$ is equivalent to one of the following: $-$, $\Box$, $\Diamond$, $\neg$, $\neg\Box$, and $\neg\Diamond$, where $-$ indicated the absence of any negation or modality.

The Modal Logic $KT4$

- The modal logic $KT4$ is also called $S4$ in the literature.
- Correspondence theory tells us that its models are precisely the Kripke models (W, R, L), where R is reflexive and transitive.
- Such structures are often very useful in computer science.
  - If Φ stands for the type of a piece of code, then $\Box\Phi$ could stand for residual code of type Φ. Thus, in the current world x this code would not have to be executed, but could be saved (= residualised) for execution at a later computation stage.
  - The formula scheme $\Box\Phi \rightarrow \Phi$, the axiom T, then means that code may be executed right away, whereas the formula scheme $\Box\Phi \rightarrow \Box\Box\Phi$ (the axiom 4) allows that residual code remain residual, i.e. we can repeatedly postpone its execution in future computation stages.
  - Such type systems have important applications in the specialisation and partial evaluation of code.

Theorem: Any sequence of modal operators and negations in $KT4$ is equivalent to one of the following: $-$, $\Box$, $\Diamond$, $\Box\Diamond$, $\Diamond\Box$, $\Box\Diamond\Box$, $\Diamond\Box\Diamond$, $\neg$, $\neg\Box$, $\neg\Diamond$, $\neg\Box\Diamond$, $\neg\Diamond\Box$, and $\neg\Diamond\Box\Diamond$.

Intuitionistic Propositional Logic

At the beginning of this course we gave a natural deduction system for propositional logic which was sound and complete with respect to semantic entailment based on truth tables. We also pointed out that the proof rules RAA, LEM and $\neg\neg$e are questionable in certain computational situations. If we disallow their usage in natural deduction proofs, we obtain a logic, called ***intuitionistic propositional logic***, together with its own proof theory. So far so good; but it is less clear what sort of semantics one could have for such a logic (again with soundness and completeness in mind).

This is where certain models of $KT4$ will do the job quite nicely. Recall that correspondence theory implies that a model $\mathcal{M} = (W, R, L)$ of $KT4$ is such that R is reflexive and transitive. The only additional requirement we impose on a model for intuitionistic propositional logic is that its labeling function L be monotone in R: xRy implies that $L(x)$ is a subset of $L(y)$. This models that atomic positive formulas persist throughout the worlds that are reachable from a given world.

Models of Intuitionistic Propositional Logic

Definition: A model of intuitionistic propositional logic is a model $\mathcal{M} = (W, R, L)$ of $KT4$ such that xRy always implies $L(x) \subseteq L(y)$. Given a propositional logic formula without negation, we define $x \Vdash \Phi$ in the usual way with the exception of the interpretation of implication and negation. For $\Phi_1 \rightarrow \Phi_2$ we define $x \Vdash \Phi_1 \rightarrow \Phi_2$ iff for all y with xRy we have $y \Vdash \Phi_2$ whenever we have we have $yh\neg \Vdash \Phi_1$. For $\neg\Phi$ we define $x \Vdash \neg\Phi$ iff for all y with xRy we have $y\neg \Vdash \Phi$.

As an example of such a model consider $W = \{x, y\}$, the relation R given by $R(x,x)$, $R(x,y)$ and $R(y,y)$. Note that R is indeed reflexive and transitive. The labelling function L satisfies $L(x) = \emptyset$ and $L(y) = \{p\}$. We claim that $x\neg \Vdash p \vee \neg p$ (recall that $p \vee \neg p$ is an instance of LEM) Clearly, we do not have $x \Vdash p$, for p is not in the set $L(x)$, which is empty. Thus, $x \Vdash p \vee \neg p$ can hold only if $x \Vdash \neg p$ holds. But $x \Vdash \neg p$ simply does not hold, since there is a world y with xRy such that $y \Vdash p$ holds, for $p \in L(y)$. Here you can see that the availability of possible worlds in the models of $KT4$ together with a 'modal interpretation' of implications and negations broke down the validity of the theorem LEM in classical logic.

Semantic Entailment

Definition: Let L be a modal logic. Such a logic is completely given by a collection of formula schemes, the axioms of L. Given a set Γ of basic modal formulas and Φ a formula of basic modal logic, we say that Γ ***semantically entails*** Φ in L and write

$$\Gamma \models_L \Phi$$

iff $\Gamma \cup L$ semantically entails Φ in basic modal logic.

Thus, we have $\Gamma \models_L \Phi$ if every Kripke model and every world x satisfying $\Gamma \cup L$ therein also satisfies Φ.

Natural Deduction

- Computing semantic entailment would be rather difficult if we had only the definition given in the previous slide. We would have to consider every Kripke model and every world in it.
- Fortunately, we have a much more usable approach, which is an extension, respectively adaptation, of the systems of natural deduction developed for propositional and predicate logic.
- We presented natural deduction proofs as linear representations of proof trees which may involve proof boxes which control the scope of assumptions, or quantifiers.
- The proof boxes have formulas and/or other boxes inside them.
- There are rules which dictate how to construct proofs. Boxes open with an assumption; when a box is closed (in accordance with a rule) we say that its assumption is discharged.
- Formulas may be repeated and brought into boxes, but may not be brought out of boxes.
- Every formula must have some justification to its right: a justification can be the name of a rule, or the word 'assumption', or an instance of the proof rule copy.

Natural Deduction Rules for Modal Logic

- We introduce a new kind of proof box, to be drawn with dashed lines. This is required for the rules for the connective $\Box$.
- Going into a dashed box means reasoning in an arbitrary related world. If at any point in a proof we have $\Box\Phi$, we could open a dashed box and put Φ in it.
- Then, we could work on this Φ, to obtain, for example, Ψ. Now we could come out of the dashed box and, since we have shown Ψ in an arbitrary related world, we may deduce $\Box\Psi$ in the world outside the dashed box.
- The rules for bringing formulas into dashed boxes and taking formulas out of them are the following:

The Rules for $KT45$

The rules $\Box$i and $\Box$e are sufficient for the modal logic K. Stronger modal logics such as $KT45$ require some extra rules if one wants to capture semantic entailment via proofs. In the case of $KT45$, this extra strength is coded up by rule forms of the axioms T, 4 and 5, as follows:

$$\frac{\Box\Phi}{\Phi}\,T \qquad \frac{\Box\Phi}{\Box\Box\Phi}\,4 \qquad \frac{\neg\Box\Phi}{\Box\neg\Box\Phi}\,5$$

Example

$\vdash_K \Box p \wedge \Box q \rightarrow \Box(p \wedge q)$.

1	$\Box p \wedge \Box q$	assumption
2	$\Box p$	$\wedge e_1$ 1
3	$\Box q$	$\wedge e_2$ 1
4	p	$\Box$e 2
5	q	$\Box$e 3
6	$p \wedge q$	$\wedge$i 4, 5
7	$\Box(p \wedge q)$	$\Box$i 4–6
8	$\Box p \wedge \Box q \rightarrow \Box(p \wedge q)$	$\rightarrow$i 1–7

Example

$\vdash_{\mathrm{KT45}} p \rightarrow \Box\Diamond p.$

1	p	assumption
2	$\Box\neg p$	assumption
3	$\neg p$	T 2
4	$\bot$	$\neg$e 1, 3
5	$\neg\Box\neg p$	$\neg$i 2–4
6	$\Box\neg\Box\neg p$	axiom 5 on line 5
7	$p \rightarrow \Box\neg\Box\neg p$	$\rightarrow$i 1–6

Example

$\vdash_{KT45} \Box\Diamond\Box p \rightarrow \Box p.$

1	$\Box\neg\Box\neg\Box p$	assumption
2	$\neg\Box\neg\Box p$	$\Box$e 1
3	$\neg\Box p$	assumption
4	$\Box\neg\Box p$	axiom 5 on line 3
5	$\bot$	$\neg$e 4, 2
6	$\neg\neg\Box p$	$\neg$i 3–5
7	$\Box p$	$\neg\neg$e 6
8	p	T 7
9	$\Box p$	$\Box$i 2–8
10	$\Box\neg\Box\neg\Box p \rightarrow \Box p$	$\rightarrow$i 1–9

Reasoning About Knowledge in a Multi-Agent System

- In a multi-agent system, different agents have different knowledge of the world.
- An agent may need to reason about its own knowledge about the world; it may also need to reason about what other agents know about the world.
- For example, in a bargaining situation, the seller of a car must consider what a buyer knows about the car's value.
- The buyer must also consider what the seller knows about what the buyer knows about the value and so on.
- Reasoning about knowledge refers to the idea that agents in a group take into account not only the facts of the world, but also the knowledge of other agents in the group.
- Example of such reasoning: ***Dean doesn't know whether Nixon knows that Dean knows that Nixon knows that McCord burgled O'Brien's office at Watergate.***

The Wise Men Puzzle

There are three wise men. It's common knowledge —known by everyone and known to be known by everyone, etc. —that there are three red hats and two white hats. The king puts a hat on each of the wise men in such a way that they are not able to see their own hat, and asks each one in turn whether they know the colour of the hat on their head. Suppose the first man says he does not know; then the second says he does not know either.

It follows that the third man must be able to say that he knows the colour of his hat. Why is this? What colour has the third man's hat?

To answer these questions, let us enumerate the seven possibilities which exist: they are

R	R	R	R	W	W	W	R	W
R	R	W	W	R	R	W	W	R
R	W	R						

where (for example) R W W refers to the situation that the first, second and third men have red, white and white hats, respectively. The eighth possibility, W W W, is ruled out by the fact that there are only two white hats.

The Wise Men Puzzle (cont'd)

Now let's think of it from the second and third men's point of view.

- When they hear the first man speak, they can rule out the possibility of the true situation being R W W, because if it were this situation, then the first man, seeing that the others were wearing white hats and knowing that there are only two white hats, would have concluded that his hat must be red.
- As he said that he did not know, the true situation cannot be R W W. Notice that the second and third men must be intelligent in order to perform this reasoning; and they must know that the first man is intelligent and truthful as well.
- In the puzzle, we assume the truthfulness and intelligence and perceptiveness of the men are common knowledge - known by everyone and known to be known by everyone, etc.

The Wise Men Puzzle (cont'd)

- When the third man hears the second man speak, he can rule out the possibility of the true situation being W R W, for similar reasons: if it were that, the second man would have said that he knew his hat was red, but he did not say this.
- The third man can also rule out the situation R R W when he hears the second man's answer, for this reason: if the second man had seen that the first was wearing red and the third white, he would have known that it must be R W W or R R W; but he would have known from the first man's answer that it couldn't be R W W, so he would have concluded it was R R W and that he was wearing a red hat; but he did not draw this conclusion, so, reasons the third man, it cannot be R R W.

The Wise Men Puzzle (cont'd)

- Having heard the first and second men speak, the third man has eliminated R W W, W R W and R R W, leaving only R R R, R W R, W R R and W W R. In all of these he is wearing a red hat, so he concludes that he must be wearing a red hat.
- Notice that the men learn a lot from hearing the other men speak. We emphasise again the importance of the assumption that they tell the truth about their state of knowledge and are perceptive and intelligent enough to come to correct conclusions.
- Indeed, it is not enough that the three men are truthful, perceptive and intelligent; they must be known to be so by the others and (in later examples) this fact must also be known, etc. Therefore, we assume that all this is common knowledge.

The Muddy Children Puzzle

There is a large group of children playing in the garden (their perceptiveness, truthfulness and intelligence is common knowledge). A certain number of children (say k) get mud on their foreheads. Each child can see the mud on others, but not on his own forehead. If $k > 1$, then each child can see another with mud on its forehead, so each one knows that at least one in the group is muddy. Consider these two scenarios:

Scenario 1. The father repeatedly asks the question 'Does any of you know whether you have mud on your own foreheadT. The first time they all answer'no'; but, unlike in the wise-men example, they don't learn anything by hearing the others answer 'no', so they go on answering 4 'no' to the father's repeated questions.

Scenario 2. The father first announces that at least one of them is muddy (which is something they know already); and then, as before, he repeatedly asks them 'Does any of you know whether you have mud on your own forehead?'. The first time they all answer 'no'. Indeed, they go on answering 'no' to the first $k-1$ repetitions of that same question; but at the k^{th} those with muddy foreheads are able to answer 'yes'.

Comments

At first sight, it seems rather puzzling that the two scenarios are different, given that the only difference in the events leading up to them is that in the second one the father announces something that they already know. It would be wrong, however, to conclude that the children learn nothing from this announcement. Although everyone knows the content of the announcement, the father's saying it makes it common knowledge among them, so now they all know that everyone else knows it, etc. This is the crucial difference between the two scenarios.

Comments (cont'd)

To understand scenario 2, consider a few cases of k.

$k = 1$ Just one child has mud. That child is immediately able to answer i yes', since she has heard the father and doesn't see any other child with mud.

$k = 2$ Say only the children a and b have mud. Everyone answers 'no' the first time. Now a thinks: since b answered 'no' the first time, he must see someone with mud. Well, the only person I can see with mud is b, so if b can see someone else it must be me. So a answers 'yes' the second time. Child b reasons symmetrically about a and also answers 'yes' the second time round.

$k = 3$ Say only the children a, b and c have mud. Everyone answers 'no' the first two times. But now a thinks: if it was just b and c with mud, they would have answered 'yes' the second time, making the argument for $k = 2$ above. So there must be a third person with mud; since I can see only b and c having mud, the third person must be me. So a answers 'yes' the third time. For symmetrical reasons, so do b and c.

And similarly for other cases of k.

The Modal Logic $KT45^n$

We now generalise the modal logic $KT45$. Instead of having just one $\Box$, it will have many, one for each agent i from a fixed set $\mathcal{A} = \{1, 2, \ldots, n\}$ of agents. We write those modal connectives as K_i (for each agent $i \in \mathcal{A}$); the K is to emphasise the application to knowledge. We assume a collection p, q, r, $\ldots$ of atomic formulas. The formula $K_i p$ means that agent i knows p; so, for example,

$$K_1 p \wedge K_1 \neg K_2 K_1 p$$

means that agent 1 knows p, but knows that agent 2 doesn't know he knows it. We also have the modal connectives E_G, where G is any subset of $\mathcal{A}$. The formula $E_G p$ means everyone in the group G knows p. If $G = \{1, 2, 3, \ldots, n\}$, then $E_G p$ is equivalent to

$$K_1 p \wedge K_2 p \wedge \cdots \wedge K_n p$$

Comments

- One might think that Φ could not be more widely known than everyone knowing it, but this is not the case. It could be, for example, that everyone knows Φ, but they might not know that they all know it.
- If Φ is supposed to be a secret, it might be that you and your friend both know it, but your friend does not know that you know it and you don't know that your friend knows it.
- Thus, $E_G E_G \Phi$ is a state of knowledge even greater than $E_G \Phi$ and $E_G E_G E_G \Phi$ is greater still.
- We say that Φ is ***common knowledge among*** G, written $C_G \Phi$, if everyone knows Φ and everyone knows that everyone knows it; and everyone knows that; and knows ***that***, etc., i.e. we may think of $C_G \Phi$ as an infinite conjunction

$$E_G \Phi \wedge E_G E_G \Phi \wedge E_G E_G E_G \Phi \wedge \cdots$$

- However, since our logics only have finite conjunctions, we cannot reduce C_G to something which is already in the logic. We have to express the infinite aspect of C_G via its semantics and retain it as an additional modal connective.
- Finally, $D_G \Phi$ means the knowledge of Φ is distributed among the group G; although no-one in G may know it, they would be able to work it out if they put their heads together and combined the information distributed among them.

Some Valid Formulas in $KT45^n$

The formula K holds for the connectives K_i, E_G, C_G and D_G, i.e. we have the corresponding formula schemes.

$$K_1\Phi \wedge K_i(\Phi \rightarrow \Psi) \rightarrow K_i\Psi$$
$$E_G\Phi \wedge E_G(\Phi \rightarrow \Psi) \rightarrow E_G\Psi$$
$$C_G\Phi \wedge C_G(\Phi \rightarrow \Psi) \rightarrow C_G\Psi$$
$$D_G\Phi \wedge D_G(\Phi \rightarrow \Psi) \rightarrow D_G\Phi$$

This means that these different 'levels' of knowledge are closed under logical consequence. For example, if certain facts are common knowledge and some other fact follows logically from them, then that fact is also common knowledge.

Some Valid Formulas in $KT45^n$ (cont'd)

Observe that E, C and D are 'box-like' connectives, in the sense that they quantify universally over certain accessibility relations. That is to say, we may define the relations R_{E_G}, R_{D_G} and R_{C_G} in terms of the relations R_i, as follows:

$$R_{E_G}(x,y) \text{ iff } R_i(x,y) \text{ for some } i \in G$$
$$R_{D_G}(x,y) \text{ iff } R_i(x,y) \text{ for all } i \in G$$
$$R_{C_G}(x,y) \text{ iff } R^k_{E_G}(x,y) \text{ for each } k \geq 1$$

It follows from this that E_G, D_G and C_G satisfy the K formula with respect to the accessibility relations R_{E_G}, R_{D_G} and R_{C_G}, respectively.

Other Valid Formulas

Since we have stipulated that the relations R_i are equivalence relations, the following formulas are valid in $KT45^n$ (for each agent i):

$K_i\Phi \rightarrow K_iK_i\Phi$	positive introspection
$\neg K_i \rightarrow K_i\neg K_i\Phi$	negative introspection
$K_i\Phi \rightarrow \Phi$	truth

These formulas also hold for D_G, since R_{D_G} is also an equivalence relation, but these don't automatically generalise for E_G and C_G. For example, $E_G\Phi \rightarrow E_GE_G\Phi$ is not valid; if it were valid, it would imply that common knowledge was nothing more than knowledge by everybody. The scheme $\neg E_G\Phi \rightarrow E_G\neg E_G\Phi$ is also not valid. The failure of these formulas to be valid can be traced to the fact that R_{E_G}, is not necessarily an equivalence relation, even though each R_i is an equivalence relation.

However, R_{E_G} is reflexive, so $E_G\Phi \rightarrow \Phi$ is valid, provided that $G \neq \emptyset$ (if $G = 0$, then $E_G\Phi$ holds vacuously, even if Φ is false).

Since R_{C_G}, is an equivalence relation, the formulas T, 4 and 5 above do hold for C_G, although the third one still requires the condition that $G \neq \Phi$.

Natural Deduction for $KT45^n$

The proof system for $KT45$ is easily extended to $KT45^n$ (but for simplicity, we omit reference to the connective D).

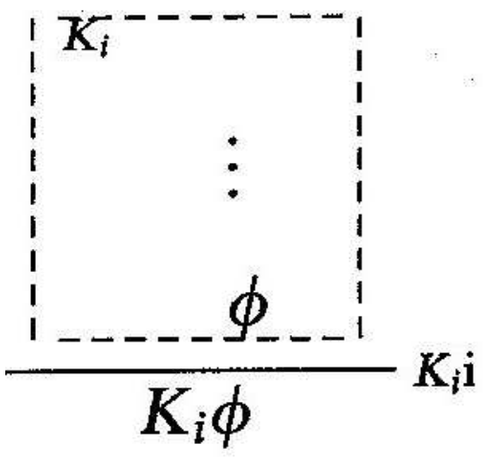
1. The dashed boxes now come in different 'flavours' for different modal connectives; we'll indicate the modality in the top left corner of the dashed box.

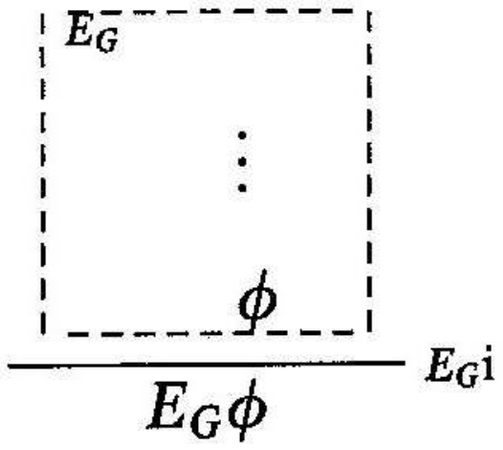
2. The axioms T, 4 and 5 can be used for any K_i, whereas axioms 4 and 5 can be used for C_G, but not for E_G.

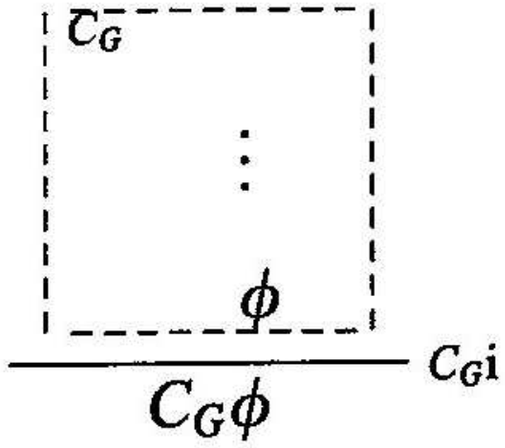
3. From $C_G\Phi$ we may deduce $E_G^k\Phi$, for any k (we call this rule CE); or we could go directly to $K_{i_1}\cdots K_{i_k}\Phi$ for any agents $i_1,\ldots,i_k$. This rule is called CK.

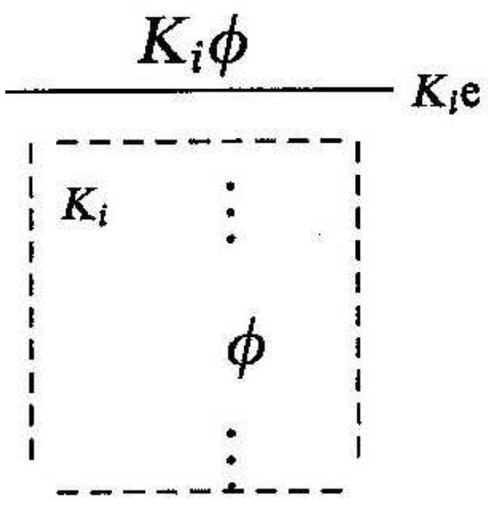
4. From $E_G\Phi$ we may deduce $K_i\Phi$ for any $i \in G$ (called EK_i). From $\bigwedge_{i\in G} K_i\Phi$ we may deduce $E_G\Phi$ (rule KE). Note that the proof rule EK_i is like a generalized and-elimination rule, whereas KE behaves like an and-introduction rule.

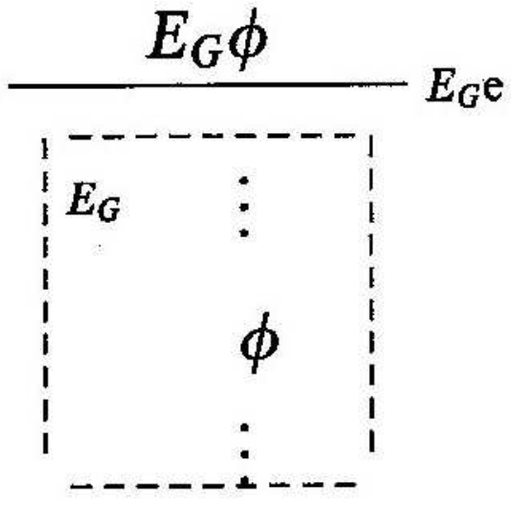
Natural Deduction Rules for $KT45^n$

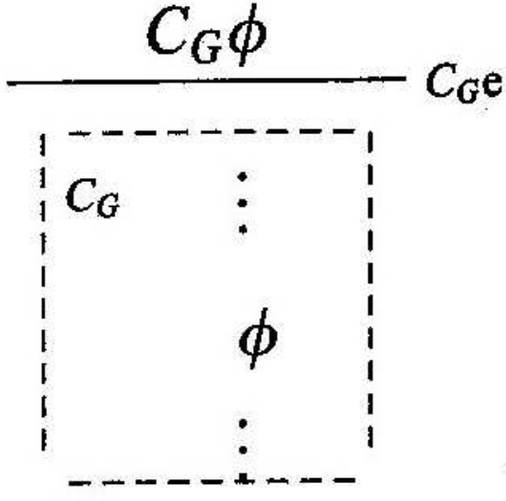












Natural Deduction Rules for $KT45^n$ (cont'd)

$$\frac{K_i\phi \text{ for each } i \in G}{E_G\phi}\ KE \qquad \frac{E_G\phi \quad i \in G}{K_i\phi}\ EK_i \qquad \frac{C_G\phi}{E_G \ldots E_G\phi}\ CE$$

$$\frac{C_G\phi}{K_{i_1} \ldots K_{i_k}\phi}\ CK \qquad \frac{C_G\phi}{C_G C_G\phi}\ C4 \qquad \frac{\neg C_G\phi}{C_G \neg C_G\phi}\ C5$$

$$\frac{K_i\phi}{\phi}\ KT \qquad \frac{K_i\phi}{K_i K_i\phi}\ K4 \qquad \frac{\neg K_i\phi}{K_i \neg K_i\phi}\ K5$$

Formalization of Wise Men Puzzle

Let p_i mean that man i has a red hat; so $\neg p_j$ means that man i has a white hat. Let Γ be the set of formulas

$$\begin{array}{ll}\{ & C(p_1 \vee p_2 \vee p_3), C(p_1 \rightarrow K_2 p_1),\ C(\neg p_1 \rightarrow K_2 \neg p_1), C(p_1 \rightarrow K_3 p_1),\ C(\neg p_1 \rightarrow K_3 \neg p_1), \\ & C(p_2 \rightarrow K_1 p_2),\ C(\neg p_2 \rightarrow K_1 \neg p_2), C(p_2 \rightarrow K_3 p_2),\ C(\neg p_2 \rightarrow K_3 \neg p_2), \\ & C(p_3 \rightarrow K_1 p_3),\ C(\neg p_3 \rightarrow K_1 \neg p_3), C(p_3 \rightarrow K_2 p_3),\ C(\neg p_3 \rightarrow K_2 \neg p_3)\ \}\end{array}$$

This corresponds to the initial set-up: it is common knowledge that one of the hats must be red and that each man can see the colour of the other men's hats. The announcement that the first man doesn't know the colour of his hat amounts to the formula

$$C(\neg K_1 p_1 \wedge \neg K_1 \neg p_1)$$

and similarly for the second man. A naive attempt at formalising the wise-men problem might go something like this: we simply prove

$$\Gamma, C(\neg K_1 p_1 \wedge \neg K_1 \neg p_1), C(\neg K_2 p_2 \wedge \neg K_2 \neg p_2) \vdash K_3 p_3,$$

i.e. if Γ is true and the announcements are made, then the third man knows his hat is red. However, this fails to capture the fact that time passes between the announcements. The fact that $C\neg K_1 p_1$ is true after the first announcement does not mean it is true after some subsequent announcement. For example, if someone announces p_1, then C_{p_1} becomes true.

Correct Formalization

The reason that this formalisation is incorrect, then, is that, although knowledge accrues with time, *lack* of knowledge does not accrue with time. If I know Φ, then (assuming that Φ doesn't change) I will know it at the next time-point; but if I do not know Φ, it may be that I do know it at the next time point, since I may acquire more knowledge.

To formalise the wise-men problem correctly, we need to break it into two entailments, one corresponding to each announcement. When the first man announces he does not know the colour of his hat, a certain positive formula Φ becomes common knowledge. Our informal reasoning explained that all men could then rule out the state R W W which, given $p_2 \vee p_2 \vee p_3$, led them to the common knowledge of $p_2 \vee p_3$. Thus, Φ is just $p_2 \vee p_3$ and we need to prove the entailments

Entailment 1: $\Gamma, C(\neg K_1 p_1 \wedge \neg K_1 \neg p_1) \vdash C(p_2 \vee p_3)$

Entailment 2: $\Gamma, C(p_2 \vee p_3), C(\neg K_2 p_2 \wedge \neg K_2 \neg p_2) \vdash K_3 p_3$.

This method requires some careful thought: given an announcement of negative information (such as a man declaring that he does not know what the colour of his hat is), we need to work out what positive knowledge formula can be derived from this and such knowledge has to be sufficient to allow us to proceed to the next round (= make even more progress towards solving the puzzle).