Properties of knowledge

(in the Kripke's semantics of possible worlds)

Can we „describe“ all formulas with modalities K_1, .., K_n, that are ever true?

Let us consider a Kripke strukture $M = (S, \pi, K_1, \ldots, K_n)$ and a formula ***A***. We define the following notions:

(i) ***A* *is valid in M*** (denoted as ***M*** **|=** ***A***), if ***A*** is true in all the states of ***M***, ie. in any state ***s*** of ***M*** holds (*M, s*) |= ***A*.**

(ii) ***A* *is satisfiable in M***, if there is a state ***t*** in ***M*** such that **(*M, t*) |=** ***A***.

(iii) ***A* *is valid*** (denoted as **|=** ***A***), if it is valid in all structures.

(iv) ***A* *is satisfiable***, if there is some structure ***M*** such that ***A*** is satisfiable in ***M***.

(v) a *formula* ***B*** *is a* ***logic consequence of A***, if ***B*** is valid in any structure ***M***, where ***A*** is valid (whenever ***M*** **|=** ***A*** , then ***M*** **|=** ***B***).

Observation. A formula *A* is **valid** (is valid in *M*) if and only if (abr. iff) the formula **¬*A*** is not **satisfiable** (is not satisfiable in *M*).

There are many valid formulas (all propositional tautologies, …)

*We search for some algorithm that would characterize all **valid formulas** and **logic consequences** using purely syntactic means (that apply transformations of formulas only)!*

Is there a FORMAL SYSTEM, that could do it?

Some examples of a FORMAL SYSTEM:

- *A set of axioms + derivation rules for propositional logic.*
- *Resolution rule for the 1st order logic.*

Let us identify some important valid formulas.

Our agents do know all the logical consequences of their knowledge:
Suppose the agent 1 knows both ***A*** and ***A*** implies ***B*** . This means that

- both formulas ***A*** and $A \to B$ are true in all the states the agent 1 considers possible,
- ***B*** must be true in all the states the agent 1 considers possible - this means that the agent 1 knows ***B***, too.

This can be written formally as: $\models (K_i A \wedge K_i (A \to B)) \to K_i B$

This formula is referred to as the **Distribution Axiom** or Kripke's axiom (denoted as **K**) because it allows to distribute the K_i operator over implication.

Agents in the Kripke's structures are very strong and competent: Let us consider a structure ***M*** *and a formula* ***A*** *valid in* ***M****. Each agent in* ***M*** *knows* ***A****.*

If A is true in all states of the structure M, then A must be true in all states of the structure M the agent considers possible. THUS:

There holds for any structure M *„if* $M \models A$, *then* $M \models K_i A$"

This observation confirms correctness of the **Knowledge Generalization** derivation rule *„If A is given, one can derive* $K_i A$ *"* for any ***i*****.**

This rule is sometimes depicted in the form $$\frac{A}{K_i A}$$

Caution!

The Generalization Rule cannot be written in the form $A \rightarrow K_i A$

This formula claims „if $\boldsymbol{A}$ is true, then the agent i knows $\boldsymbol{A}$ “.
But this is NOT a **valid formula**!

An agent does not have to know all facts that are true in the considered state:

In the case of muddy children a child with muddy forehead does not know this fact first. This knowledge is acquired later!

Our agents know all formulas valid in the considered structure, but nothing more! In other words they know only those formulas that are **true** *necessarily*.

They do not know formulas, that happen to be true in some of the worlds only (*by chance*).

Our agent does not have to know all facts that are true. But *if the agent knows something, then it holds*:

$$\models K_i A \rightarrow A$$

This property is often referred to as the ***Knowledge Axiom*** or the ***Truth Axiom*** (denoted as **T**).

Validity of this axiom is a consequence of reflexivity of the admissibility relation describing what the agent considers possible:

If $\boldsymbol{K_i A}$ is true in some world (M, s), $\boldsymbol{A}$ must hold in all states the agent $\boldsymbol{i}$ considers possible – this includes (M, s), since the considered admissibility relation is assumed to be reflexive.

{Philosophers use this axiom to highlight the difference between **knowledge** and **belief**.}

In the case we want to describe belief of an agent, not its knowledge, it is necessary to replace the Truth Axiom

$$\models K_i A \rightarrow A$$

by a weaker requirement that ensures consistency: $\neg K_i false$

This is the ***Consistency Axiom*** , often referred to as **D**.

The next two properties describe what the agents know about their knowledge thanks to introspection. *Our agents know, what they know and what they do not know:*

$$\models K_i A \to K_i K_i A$$

$$\models \neg K_i A \to K_i \neg K_i A$$

The first property is called ***Positive Introspection Axiom*** (often denoted as **4**),

The second one is the ***Negative Introspection Axiom*** (often denoted as **5**).

Both are valid in the Kripke structures where admissibily relations are equivalences. ***Try to prove it!***

Formal (axiomatic) system K_n

Axioms: **A1.** All the propositional tautologies

A2. $(K_i \alpha \wedge K_i(\alpha \rightarrow \beta)) \rightarrow K_i(\beta)$

Derivation rules:

R1. From the formulas α and $\alpha \rightarrow \beta$ derive β (**Modus Ponens**)

R2. From the formula α derive $K_i \alpha$ (**Knowledge Generalization Rule**)

Proof of a formula φ in a formal system is a sequence of formulas $\delta_1, \delta_2, \dots, \delta_n$ such that δ_n is the formula φ and for any δ_i ($i < n+1$) holds

- either δ_i is an axiom of the considered formal system
- or there are numbers j and k smaller than i such that δ_i is the result of derivation rule application on δ_j or on δ_j are δ_k .

The formula φ is **provable in the formal system** (denoted as $\vdash \varphi$), if φ has a proof in this system.

Properties of the formal system K_n

Axioms:

A1. All the propositional tautologies

A2. $(K_i \alpha \wedge K_i (\alpha \to \beta)) \to K_i (\beta)$

Derivation rules:

R1. From the formulas α and $\alpha \to \beta$ derive β (**Modus Ponens**)

R2. From the formula α derive $K_i \alpha$ (**Knowledge Generalisation**)

What is the **relation between**

- **the formulas** that are **provable in the** system K_n and
- the formulas **valid** in all the Kripke structures with ***n*** agents ?

Formal system is **correct**, if any provable formula is also valid (ie. „For any formula A there holds that if $\vdash$ A than $\models A$ “).

Formal system is **complete**, if all valid formulas can be proven (ie. „For any formula A hold that if $\models A$ than $\vdash A$“)

$\mathbf{K}_n \vdash K_i(\alpha \wedge \beta) \rightarrow K_i \alpha$:

Formal proof – sequence of formulas**:** [Each formula in the sequence ***must provide a reference to one of*** $\mathbf{K}_n$ ***axioms*** or a ***precise description of the derivation rule*** *as it is applied* to the formulas appearing earlier in the proof].

1. $(\alpha \wedge \beta) \rightarrow \alpha$ [Prop.tautology]

2. $K_i((\alpha \wedge \beta) \rightarrow \alpha)$ [**KG:** 1, ie. "**KG** is applied to the formula from the row 1]

3. $(K_i(\alpha \wedge \beta) \wedge K_i((\alpha \wedge \beta) \rightarrow \alpha)) \rightarrow K_i \alpha$ [$\mathbf{K}_n$ distribution axiom]

4. $((K_i(\alpha \wedge \beta) \wedge K_i((\alpha \wedge \beta) \rightarrow \alpha)) \rightarrow K_i \alpha)$

$\rightarrow (K_i((\alpha \wedge \beta) \rightarrow \alpha) \rightarrow (K_i(\alpha \wedge \beta) \rightarrow K_i \alpha))$

[Prop. tautology $((p \wedge q) \rightarrow r) \rightarrow (q \rightarrow (p \rightarrow r))$]

5. $K_i((\alpha \wedge \beta) \rightarrow \alpha) \rightarrow (K_i(\alpha \wedge \beta) \rightarrow K_i \alpha)$ [**MP:** 3,4]

6. $K_i(\alpha \wedge \beta) \rightarrow K_i \alpha$ [**MP:** 2,5]

1. $\mathbf{K}_n \vdash K_i(\alpha \wedge \beta) \rightarrow K_i\alpha$ [see the former page]
2. $\mathbf{K}_n \vdash K_i(\alpha \wedge \beta) \rightarrow K_i\beta$ [This proof is a minor modification of that of the formula on the line 1]
3. $(K_i(\alpha \wedge \beta) \rightarrow K_i\alpha) \rightarrow ((K_i(\alpha \wedge \beta) \rightarrow K_i\beta) \rightarrow (K_i(\alpha \wedge \beta) \rightarrow (K_i\alpha \wedge K_i\beta)))$
   $[(\rho \rightarrow \varphi) \rightarrow ((\rho \rightarrow \psi) \rightarrow (\rho \rightarrow (\varphi \wedge \psi)))$ [propositional tautology]
4. $(K_i(\alpha \wedge \beta) \rightarrow K_i\beta) \rightarrow (K_i(\alpha \wedge \beta) \rightarrow (K_i\alpha \wedge K_i\beta))$ [**MP**: 1,3]
5. $K_i(\alpha \wedge \beta) \rightarrow (K_i\alpha \wedge K_i\beta)$ [**MP**: 2,4]

Claim 1: $\mathbf{K}_n \vdash K_i(\alpha \wedge \beta) \equiv K_i\alpha \wedge K_i\beta$

Proof:

- The implication $\rightarrow$ has been proven above.
- The inverse implication is on the next page.

$\mathbf{K}_n \vdash K_i\alpha \wedge K_i\beta \rightarrow K_i(\alpha \wedge \beta)$

6.	$\alpha \rightarrow (\beta \rightarrow (\alpha \wedge \beta))$	[prop.tautology]
7.	$K_i(\alpha \rightarrow (\beta \rightarrow (\alpha \wedge \beta)))$	[**KG**:6]
8.	$K_i\alpha \rightarrow (K_i(\alpha \rightarrow (\beta \rightarrow (\alpha \wedge \beta))) \rightarrow K_i(\beta \rightarrow (\alpha \wedge \beta)))$	[distribution axiom A2]
9.	$K_i(\alpha \rightarrow (\beta \rightarrow (\alpha \wedge \beta))) \rightarrow (K_i\alpha \rightarrow K_i(\beta \rightarrow (\alpha \wedge \beta)))$	[„prop.modification of“ 8]
10.	$(K_i\alpha \rightarrow K_i(\beta \rightarrow (\alpha \wedge \beta)))$	[**MP**: 9,7]
11.	$(K_i\alpha \rightarrow K_i(\beta \rightarrow (\alpha \wedge \beta))) \rightarrow (K_i\beta \rightarrow (K_i\alpha \rightarrow K_i(\beta \rightarrow (\alpha \wedge \beta))))$	[Prop.tautology]
12.	$K_i\beta \rightarrow (K_i\alpha \rightarrow K_i(\beta \rightarrow (\alpha \wedge \beta)))$	[**MP** 11,10]
13.	$(K_i\alpha \wedge K_i\beta) \rightarrow K_i(\beta \rightarrow (\alpha \wedge \beta))$	[„prop.modification of“ 12]
14.	$(K_i\alpha \wedge K_i\beta) \rightarrow (K_i\beta \wedge K_i(\beta \rightarrow (\alpha \wedge \beta)))$	[„prop.modification of“13, see *]
15.	$K_i\beta \wedge K_i(\beta \rightarrow (\alpha \wedge \beta)) \rightarrow K_i(\alpha \wedge \beta)$	[A2]
16.	$((K_i\alpha \wedge K_i\beta) \rightarrow (K_i\beta \wedge K_i(\beta \rightarrow (\alpha \wedge \beta)) \rightarrow K_i(\alpha \wedge \beta)))$ $\rightarrow (((K_i\alpha \wedge K_i\beta) \rightarrow K_i\beta \wedge K_i(\beta \rightarrow (\alpha \wedge \beta))) \rightarrow ((K_i\alpha \wedge K_i\beta) \rightarrow K_i(\alpha \wedge \beta)))$	[Prop.taut. * *]
17.	$(K_i\beta \wedge K_i(\beta \rightarrow (\alpha \wedge \beta)) \rightarrow K_i(\alpha \wedge \beta)) \rightarrow ((K_i\alpha \wedge K_i\beta) \rightarrow (K_i\beta \wedge K_i(\beta \rightarrow (\alpha \wedge \beta)) \rightarrow K_i(\alpha \wedge \beta)))$	[P.t. * * *]
18.	$((K_i\alpha \wedge K_i\beta) \rightarrow (K_i\beta \wedge K_i(\beta \rightarrow (\alpha \wedge \beta)) \rightarrow K_i(\alpha \wedge \beta)))$	[**MP**: 15,17]
19.	$(((K_i\alpha \wedge K_i\beta) \rightarrow K_i\beta \wedge K_i(\beta \rightarrow (\alpha \wedge \beta))) \rightarrow ((K_i\alpha \wedge K_i\beta) \rightarrow K_i(\alpha \wedge \beta)))$	[**MP**: 16, 18]
20.	$(K_i\alpha \wedge K_i\beta) \rightarrow K_i(\alpha \wedge \beta)$	[**MP**: 19, 14]

* If |-(A&B)→C then |- (A&B)→(B & C))

* * (α→ (β→γ)) → ((α → β) → (α → γ))

* * * (α→ (β→ α))

Theorem (see the lab work presentation).

For all structures $\boldsymbol{M}$ with n agents where the admissibility relations are interpreted by relations that are equivalences, there holds for any formulas A, B :

$$(i)\ \ M \models (K_i A \wedge K_i(A \to B)) \to K_i B$$

$$(ii)\ \ if \quad M \models A \quad than \quad M \models K_i A$$

$$(iii)\ \ M \models K_i A \to A$$ reflexivity

$$(iv)\ \ M \models K_i A \to K_i K_i A$$ transitivity

$$(v)\ \ M \models \neg K_i A \to K_i \neg K_i A$$ symmetric+transitive r.

Axioms of propositional modal logics

1. **Propositional tautologies**
2. **Distribution Axiom** (denoted as **K**) $(K_i A \wedge K_i(A \to B)) \to K_i B$
3. **Knowledge Axiom)** (denoted as **T**) $K_i A \to A$ r
4. **Positive Introspection Axiom** (den.as **4**) $K_i A \to K_i K_i A$ t
5. **Negative Introspection Axiom** (den.as **5**) $\neg K_i A \to K_i \neg K_i A$ s+t
6. **Consistency Axiom** (den.as **D**) $\neg K_i \mathit{false}$

Derivation rules:

R1. From the formulas α and $\alpha \to \beta$ derive β (**Modus Ponens**)

R2. From the formula α derive $K_i\, \alpha$ (**Knowledge Generalization**)

Proof of a formula φ in the formal system **under assumption** α is a sequence of formulas $\delta_1, \delta_2, \ldots, \delta_n$ such that δ_n is the formula φ and for any δ_i ($i < n+1$) holds

- either δ_i is an axiom of the considered formal system or the assumption α
- or there are numbers j and k smaller than i such that δ_i is the result of derivation rule application on δ_j or on δ_j are δ_k .

The formula φ is **provable** in the formal system **under assumption** α (denoted as $\alpha \vdash \varphi$), if φ has a proof **under assumption** α .

Claim 1 $\mathbf{K}_n$, $(\varphi \rightarrow \psi) \vdash K_i\,\varphi \rightarrow K_i\,\psi$

(The formula $K_i\,\varphi \rightarrow K_i\,\psi$ is a consequence of the assumption $(\varphi \rightarrow \psi)$ in the formal system $\mathbf{K}_n$) : $(\varphi \rightarrow \psi)$ [assumption]

1. $K_i(\varphi \rightarrow \psi)$ [KG „assumption“]
2. $K_i\varphi \rightarrow (K_i(\varphi \rightarrow \psi) \rightarrow K_i\psi)$ [distribution axiom **K**]
3. $(K_i\varphi \rightarrow (K_i(\varphi \rightarrow \psi) \rightarrow K_i\psi)) \rightarrow (K_i(\varphi \rightarrow \psi) \rightarrow (K_i\varphi \rightarrow K_i\psi))$
   [Prop-T1: $(\varphi \rightarrow (\psi \rightarrow \tau)) \rightarrow (\psi \rightarrow (\varphi \rightarrow \tau))$]
4. $K_i(\varphi \rightarrow \psi) \rightarrow (K_i\varphi \rightarrow K_i\psi)$ [MP 2,3]
5. $(K_i\varphi \rightarrow K_i\psi)$ [MP 1,4]

Claim 2: Let the formulas φ , ψ be equivalent (ie. the formula $(\varphi \rightarrow \psi) \wedge (\psi \rightarrow \varphi)$ is a tautology, denoted as $\varphi \equiv \psi$). There holds $\mathbf{K}_n$, $\varphi \equiv \psi \mid\!- K_i\,\varphi \equiv K_i\,\psi$.

This is a direct consequence of the above statement .

Let us denote by $\boldsymbol{\mathcal{M}}_n(\boldsymbol{\Phi})$ the set of all Kripke structures over the set $\boldsymbol{\Phi}$ of primitive propositions and a set of $\boldsymbol{n}$ agents. Denote that no requirements are set on the relations $\boldsymbol{K_i}$ in this case.

Let $\boldsymbol{\mathcal{M}}_n^{rst}(\boldsymbol{\Phi})$ be the subset of $\boldsymbol{\mathcal{M}}_n(\boldsymbol{\Phi})$ consisting of all the Kripke structures from where all the admissibility relations have the identified properties ***rst***, namely they are:

reflexive

symetric

transitive.

(They are **equivalences**).

Theorem 1: The system $\mathbf{K}_n$ represents correct and complete syntactic description of all formulas that are valid in the set $\mathcal{M}_n(\Phi)$ of all Kripke struktures ($\mathbf{K}_n$ is an axiomatization w.r.t. $\mathcal{M}_n(\Phi)$).

Theorem 2:

Let T be the axiom $K_i A \rightarrow A$. The system $\mathbf{T}_n = (\mathbf{K}_n + \text{axiom T})$ is the axiomatization w.r.t. $\mathcal{M}_n^{r}(\Phi)$.

Let 4 be the axiom $K_i A \rightarrow K_i K_i A$. The system $\mathbf{S}4_n = (\mathbf{T}_n + \text{axiom 4})$) is the axiomatization w.r.t. $\mathcal{M}_n^{rt}(\Phi)$.

Let 5 be the axiom $\neg K_i A \rightarrow K_i \neg K_i A$. The systém $\mathbf{S}5_n = (\mathbf{S}4_n +$ axiom 5) is the axiomatization w.r.t. k $\mathcal{M}_n^{rts}(\Phi)$.

.

Some more valid statements:

c1) K2, T(Axiom 3) $\vdash \neg K_i \mathit{false}$

1. $K_i \mathit{false} \rightarrow \mathit{false}$ [A3]
2. $\neg \mathit{false} \rightarrow (\neg K_i \mathit{false})$ [prop.modification of 1]
3. $\neg \mathit{false}$ [prop.tautology]
4. $\neg K_i \mathit{false}$ [MP: 3,2]

c2) K2, T $\vdash \neg K_i \alpha \vee \neg K_i \neg K_i \alpha$

c3) K2, T $\vdash \neg K_i (\alpha \wedge \neg K_i \alpha)$

1. $K_i \neg K_i \alpha \rightarrow \neg K_i \alpha$ [A3, Truth Axiom]
2. $\neg K_i \neg K_i \alpha \vee \neg K_i \alpha$ [prop.modification of $\rightarrow$ in 1], *viz* a1
3. $\neg (K_i \neg K_i \alpha \wedge K_i \alpha)$ [prop.modification of $\vee$ in 2]
4. $\neg K_i (\neg K_i \alpha \wedge \alpha)$ [transitivity of K_i in the formula 3], *viz* a2

Some more relations that can be proven:

a) $(\mathrm{K}_n + \mathrm{A6}) \vdash \neg (K_i \boldsymbol{\alpha} \wedge K_i \neg \boldsymbol{\alpha})$

b) $(\mathrm{K}_n + \mathrm{A3}) \vdash \mathrm{A6}$

c) $\mathbf{K}_n \vdash K_i \neg (\boldsymbol{p} \rightarrow K_i \boldsymbol{p}) \equiv K_i (\boldsymbol{p} \wedge \neg K_i \boldsymbol{p}) \equiv (K_i \boldsymbol{p} \wedge K_i (\neg K_i \boldsymbol{p}))$

d) It is not possible to prove $K_i \neg (\boldsymbol{p} \rightarrow K_i \boldsymbol{p})$ in $(\mathrm{K}_n + \mathrm{A3})$.

$E_G \quad C_G \quad D_G$

Let G be a subset of {1, 2, . . . , n}, $E_G A$ holds iff every agent from G knows A. Thus

Axiom C1. $E_G A \Leftrightarrow \underset{i \in G}{\&} K_i A$

Intuitively, **common knowledge** specifies something „*what is crystal clear to everyone*“. It should be no surprise that common knowledge has the properties that have been described in the **Distribution Axiom**, in the **Knowledge Axiom**, and in the **Positive** and **Negative Introspection Axioms**, see the next page.

Common knowledge of two groups of agents:

$$\mathit{If} \quad Q \subseteq G \quad \mathit{then} \quad C_G A \rightarrow C_Q A$$

It can be verified that the following formulas are valid (they are true in all Kripke structures):

$$(i) \quad (C_G A \,\&\, C_G(A \to B)) \to C_G B$$

$$(ii) \quad C_G A \to A$$

$$(iii) \quad C_G A \to C_G C_G A$$

$$(iv) \quad \neg C_G A \to C_G \neg C_G A$$

The assumptions on properties of the underlying admissibility relations for all $\boldsymbol{K_i}$ are the same as in the case of reasoning about knowledge.

Distributed knowledge

charakterize knowledge the agents can acquire when „*all of them share all their individual knowledge*“.

Even this modal operator has simmilar properties (axioms) as knowledge of a single agent. Let us point to some specific cases:

- *Distributed knowledge in the group with a single agent is that of the agent*, namely $\models D_{\{i\}} A \leftrightarrow K_i A$
- *The bigger the considered group the bigger their distributed knowledge :*

$$If \quad G \subseteq Q \; then \; \models D_G A \rightarrow D_Q A$$

Task6a Could the modality be defined as a boolean function?(1 point)

Let us consider for simplicity only Kripke structures with a single agent whose knowledge is described by the modal operator ***K***. We know that in all the corresponding Kripke structures where ***K*** is interpreted by equivalence there holds for any formula α

a) the formula ***K*** $\alpha \rightarrow \alpha$ (Knowledge Axiom) is valid ,

b) but the formulas $\alpha \rightarrow$ ***K*** α and $\neg$ ***K*** α are not valid.

Utilize these facts to show that such a behaviour of the modal operator ***K*** cannot be encoded by any boolean function (ie. Truth values defined by a table).

Hint: Suppose the truth value of ***K*** α can be calculated from the truth value of α using a truth table for ***K*** (in the same way as $\neg\ \alpha$ is calculated form α). Consider all possible truth tables for ***K*** and show that none of them grants the properties a) and b) mentioned above.

Task 6b Ann and Bob (2 points)

Ann and **B**ob take part in a quizz. First, the organizer selects from an urn a natural number ***n*** < **10**, that he writes on the forehead of one of the players and continues by writing the neighboring number (either ***n*+1** or ***n*-1**) on the forhead of the second player. Neither **A**nn nor **B**ob knows her/his number – each sees only the other´s forehead. They can take turns in announcing nothing but „*I do not know my number*.“ or „*I know my number*.“ Suppose Ann starts and she can see the symbol **5**.

- Who will be the first to identify her/his number?
- Demonstrate your conclusion about the winner using the corresponding Kripke structure and its modification during information exchange between **A** and **B**.