Jaké vlastnosti mají znalosti

(v sémantice možných světů)

Jeden ze způsobů jak charakterizovat naši interpretaci znalostí je popsat formule, které jsou vždy pravdivé.

Je-li dána struktura

$$M = (S, \pi, K_1, \ldots, K_n)$$

(i) říkáme, že *formule* ***A*** *je pravdivá (validní) v* ***M*** a píšeme $\boldsymbol{M} \models \boldsymbol{A}$, jestliže $(M, s) \models \boldsymbol{A}$ v každém stavu ***s***.
(ii) říkáme, že *formule* ***A*** *je splnitelná v* *M* , jestliže $(\boldsymbol{M}, \boldsymbol{t}) \models \boldsymbol{A}$ v nějakém stavu ***t*** .
(iii) říkáme, že ***formule A je validní*** a píšeme $\models \boldsymbol{A}$, je-li pravdivá (validní) ve všech strukturách.
(iv) říkáme, že *formule* ***A*** *je splnitelná*, je-li splnitelná v nějaké struktuře.

Platí

formule ***A*** je validní (validní v *M*) právě když formule $\neg\boldsymbol{A}$ není splnitelná (není splnitelná v *M*).

Připomeňme, že stále předpokládáme, že relace K_i jsou ekvivalence.

Validních formulí je jistě velké množství (všechny výrokové tautologie, …)

Hledáme nějaký způsob, jak je charakterizovat pomocí syntaktických prostředků.

Existuje FORMÁLMÍ SYSTÉM, který to dokáže?

Při zavedeném pojetí znalostí platí, je že agent zná všechny logické důsledky svých znalostí.

Ví-li agent *A* a také ví, že *A* implikuje *B*, potom obě formule *A* a $A \rightarrow B$ jsou pravdivé ve všech světech, které agent považuje za možné,

tedy také *B* musí být pravdivé ve všech světech, které agent považuje za možné, takže musí znát i *B* .

Odtud dostáváme $\models (K_i A \wedge K_i(A \rightarrow B)) \rightarrow K_i B$

Této formuli se říká **axiom distribuce** nebo také Kripkův axiom a označuje se jako **K**, protože dovoluje distribuovat operátor K_i přes implikaci.

Druhá vlastnost našeho pojetí znalosti zaručuje, že máme dostatečně silné agenty a pramení z toho, že agent zná všechny validní formule v dané struktuře.

Je-li ***A*** pravdivá ve všech stavech (tj. možných světech) struktury ***M*** , pak je pravdivá ve všech světech, které agent v daném světě považuje za možné. ⊥

Pro libovolnou strukturu ***M*** *platí*

je-li $\boldsymbol{M} \models \boldsymbol{A}$, *pak* $\boldsymbol{M} \models \boldsymbol{K_i A}$

Tak získáváme ***Pravidlo generalizace znalostí :*** $\dfrac{\vdash A}{\vdash K_i A}$

$\vdash$

Pozor! Pravidlo generalizace je něco jiného než implikace

$$A \rightarrow K_i A$$

která tvrdí „je-li *A* pravdivá, potom agent *i* to ví“: toto NENÍ validní formule! Agent nemusí vědět všechny věci, které jsou pravdivé.
{*Například při hře zablácených dětí může mít jedno z nich zablácené čelo, ale nemusí to vědět.*}

Agenti znají všechny validní formule. Jinými slovy znají JEN formule, které jsou *nutně* pravdivé.

To však neplatí o formulích, které jsou (*řízením osudu*) pravdivé v nějakém možném světě.

Ačkoliv agent nemusí znát všechna fakta, která jsou pravdivá, ale když agent něco ví, pak to platí:

$$\models K_i A \rightarrow A$$

Tato vlastnost se obvykle nazývá ***Axiom znalostí*** nebo ***Axiom pravdy*** a často se označuje **T**.

Odůvodnění axiomu: svět, ve kterém se agent nachází je vždy takový, který pokládá za možný. Platí-li $K_i A$ v nějakém světě (M, s), pak A platí ve všech světech, které i považuje za možné, tedy i v (M, s).

{Filosofové tento axiom považují za hlavní odlišení mezi **znalostí** a **přesvědčením** (vírou, belief). Mohu mít nepravdivé přesvědčení, ale nemohu vědět něco, co není pravda!}

V případě, že chceme popisovat to, čemu agent věří, nikoliv to, co ví, nahradíme axiom pravdy

$$\models K_i A \rightarrow A$$

slabším požadavkem:

Axiom konzistence požaduje $\neg K_i$ *false*, který se často se označuje **D**.

Zdůvodnění:

$K_i A$ v nějakém světě (M, s), pak A platí ve všech světech, které i považuje za možné, tedy i v (M, s).

{Filosofové tento axiom považují za hlavní odlišení mezi **znalostí** a **přesvědčením** (vírou, belief). Mohu mít nepravdivé přesvědčení, ale nemohu vědět něco, co není pravda!}

Další dvě vlastnosti se týkají přirozeného požadavku, aby agenti věděli o svých znalostech (pomocí introspekce).

Agenti vědí, co vědí a vědí, co nevědí.

$$\models K_i A \rightarrow K_i K_i A$$
$$\models \neg K_i A \rightarrow K_i \neg K_i A$$

První z těchto vlastností se jmenuje ***Axiom pozitivní introspekce*** (často označovaný **4**),

druhá ***Axiom negativní introspekce*** (často označovaný (často označovaný **5**).

Označme $\mathcal{M}_n(\Phi)$ množinu všech Kripkeho struktur nad množinou prvotních formulí **Φ** pro ***n*** agentů takových, že na relace K_i nejsou kladeny žádné požadavky.

$\mathcal{M}_n^{rst}(\Phi)$ nechť je pak podmnožina $\mathcal{M}_n(\Phi)$, ve které relace přípustnosti splňují vlastnosti ***rst***.

Věta.

Pro všechny struktury M pro n agentů s relacemi K_i, které jsou ekvivalencemi, a pro libovolné formule A, B platí

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

$$(ii)\ \ je-li\ \ M \models A\ \ \ potom\ \ M \models K_i A$$

$$(iii)\ \ M \models K_i A \rightarrow A$$

$$(iv)\ \ M \models K_i A \rightarrow K_i K_i A$$

$$(v)\ \ M \models \neg K_i A \rightarrow K_i \neg K_i A$$

V Kripkeho strukturách jsme ověřili, že validní jsou

- **Distribuční axiom** (označovaný někdy jako **K**)
- **Pravidlo generalizace znalostí**
- **Axiom znalostí (Axiom pravdy)** (ozn. jako **T**)
- **Axiom pozitivní introspekce** (ozn. jako **4**)
- **Axiom negativní introspekce** (ozn.jako **5**)
- **Axiom konzistence** (ozn.jako **D**)

•Distribuční axiom **K** a **pravidlo generalizace znalostí** jsme ověřili bez jakýchkoli předpokladů o relacích K_i,

•validnost axiomu znalostí **T** plyne z reflexivity,

•validnost axiomu pozitivní introspekce **4** plyne z tranzitivity,

•a validnost axiomu negativní introspekce **5** plyne ze symetrie a tranzitivity relací K_i.

Formální (axiomatický) systém K_n

Axiomy: **A1.** Všechny tautologie výrokové logiky

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

Odvozovací pravidla:

R1. Z α formulí $\alpha \Rightarrow \beta$ odvoďte β (modus ponens)

R2. Z formule α odvoď $K_i \alpha$ (pravidlo zobecnění znalosti)

V K_n lze dokázat například tvrzení $K_i(\alpha \wedge \beta) \Rightarrow K_i \alpha$ (viz dále).

Značení **pro dokazatelnost:** $K_n \vdash K_i(\alpha \wedge \beta) \Rightarrow K_i \alpha$

Věta:

Systém K_n představuje korektní a úplný syntaktický popis všech formulí validních v množině $\mathcal{M}_n(\Phi)$ Kripkeho struktur.

$K_n \vdash K_i(\alpha \wedge \beta) \Rightarrow K_i \alpha$:

1. $(\alpha \wedge \beta) \Rightarrow \alpha$ [A1]

2. $K_i((\alpha \wedge \beta) \Rightarrow \alpha)$ [1, **R2**]

3. $(K_i(\alpha \wedge \beta) \wedge K_i((\alpha \wedge \beta) \Rightarrow \alpha)) \Rightarrow K_i \alpha$ [A2]

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

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

[**A1**, jde o instanci tautologie $((p \wedge q) \Rightarrow r) \Rightarrow (q \Rightarrow (p \Rightarrow r))$]

5. $K_i((\alpha \wedge \beta) \Rightarrow \alpha) \Rightarrow (K_i(\alpha \wedge \beta) \Rightarrow K_i \alpha)$ [3,4, **R1**]

6. $K_i(\alpha \wedge \beta) \Rightarrow K_i \alpha$ [2,5, **R1**]

Věta 1: Systém K_n představuje korektní a úplný syntaktický popis všech formulí validních v množině $\mathcal{M}_n(\Phi)$ Kripkeho struktur (K_n je axiomatizací vzhledem k $\mathcal{M}_n(\Phi)$).

Věta 2:

Systém T_n = (K_n + axiom T) je axiom. vzhledem k $\mathcal{M}_n^{r}(\Phi)$.

Systém $S4_n$ = (T_n + axiom 4) je axiom. vzhledem k $\mathcal{M}_n^{rt}(\Phi)$.

Systém $S5_n$ = ($S4_n$ + axiom 5) je axiom. vzhledem k $\mathcal{M}_n^{rts}(\Phi)$.

.

Následující vztahy lze dokázat (viz cvičení):

- (K_n + A6) $\vdash \neg(K_i \alpha \wedge K_i \neg \alpha)$.
- $\neg K_i(\alpha \wedge K_i \neg \alpha) \equiv \neg K_i \alpha \vee \neg K_i \neg K_i \alpha$
- $K_i(\alpha \wedge \beta) \equiv K_i \alpha \wedge K_i \beta$.

$\boxed{E_G \quad C_G \quad D_G}$

G je podmnožinou {1, 2, . . . , n}, $E_G A$ je **pravdivé**, právě když každý agent z G ví A. Tedy

$$\models E_G A \quad \Leftrightarrow \quad \bigcap_{i \in G} K_i A$$

Intuitivně, **společná** (všeobecná) **znalost** je to „*co je každému zcela jasné a ví to*“, proto nepřekvapí, že společná znalost má všechny vlastnosti znalostí podobné **Distribučnímu axiomu**, **Axiomu znalostí**, **Axiomům pozitivní** a **negativní introspekce**.

(Cvičení)

Pro všeobecnou znalost mezi skupinami agentů platí

$$Je-li \quad G' \subseteq G \quad potom \quad C_G A \rightarrow C_{G'} A$$

Cvičení. Dokažte, že následující formule jsou validní:

(i) $(C_G A \wedge C_G(A \rightarrow B)) \rightarrow C_G B$

(ii) $C_G A \rightarrow A$

(iii) $C_G A \rightarrow C_G C_G A$

(iv) $\neg C_G A \rightarrow C_G \neg C_G A$

Předpoklady o relacích K_i jsou stejné, jako při odvození odpovídajících axiomů.

Axiom o pevném bodu.

Nechť p je výrok „alespoň jeden z vás má zablácené čelo“.

Připomeňme, že ve hře zablácených dětí, se z p stala všeobecná znalost v okamžiku, kdy otec řekl p a všechny děti věděly dvě věci: že platí p a že jsou v takové situaci.

Axiom.

$$\models C_G A \leftrightarrow E_G(A \wedge C_G A)$$

{ $C_G A$ je pevným bodem funkce $f(x) = E_G(A \wedge x)$ }

Následující odvozovací pravidlo dává způsob, jak ukázat, že v nějaké struktuře platí společná znalost.

Indukční pravidlo

Pro všechny struktury M platí

je - li $M \models A \to E_G(B \wedge A)$, potom $M \models A \to C_G B$

Idea. Předpoklad indukčního pravidla dává možnost indukcí podle k dokázat, že formule

$A \to E^k(B \wedge A)$ je validní pro každé k.

Následující tvrzení dává sémantické odůvodnění charakteristiky operátoru E_G, Axiomu o pevném bodu a Indukčního pravidla.

Věta - shrnutí

Pro všechny struktury M, všechny neprázdné podmnožiny G množiny {1, 2, . . . , n} a libovolné formule A, B platí

(i) $M \models E_G A \Leftrightarrow \bigcap_{i \in G} K_i A$

(ii) $M \models C_G A \leftrightarrow E_G(A \wedge C_G A)$

(iii) $je - li\ M \models A \to E_G(B \wedge A),\ potom\ M \models A \to C_G B$

Důkaz. (nepředpokládá, že relace K jsou ekvivalence)

(i) $M \models E_G A \Leftrightarrow \bigcap_{i \in G} K_i A$

přímo plyne ze sémantiky E_G

(ii) $M \models C_G A \leftrightarrow E_G(A \wedge C_G A)$

Připomeňme

$(M,s) \models C_G A$ *právě když* $(M,t) \models A$

pro všechny stavy t, které jsou G-dosažitelné z s.

$(ii) \to$ Předpokládejme

$(M,s) \models C_G A$

potom

$(M,t) \models A$

pro všechny stavy t, G-dosažitelné z s.

Rozdělme cestu na dvě části, nechť u je G-dosažitelné z s v jednom kroku a t je G-dosažitelné z u.

Potom $(M,u) \models A$ a $(M,t) \models A$

takže $(M,u) \models A \wedge C_G A$

pro všechny stavy u, G-dosažitelné z s v jednom kroku.

Odtud $(M,s) \models E_G(A \wedge C_G A)$ podle (i).

$(ii) \leftarrow$ Předpokládejme

$(M,s) \models E_G(A \wedge C_G A)$ (1)

Nechť t je G-dosažitelné z s a s' je první uzel na G-cestě z s do t. Z (1) plyne

$(M,s') \models A \wedge C_G A$ (2)

Uvažujeme dva případy s' je t nebo t je G-dosažitelné z s'.

V prvním případě je A pravdivé v t, protože t je s' a v s' A platí. Ve druhém případě je A pravdivé v t, protože t je G-dosažitelné z s' a z (2) plyne

$(M,s') \models C_G A.$

Protože A je pravdivé v každém stavu t, G-dosažitelném z s, dostáváme $(M,s) \models C_G A.$

(iii) $je-li\ \ M \models A \to E_G(B \wedge A),\ \ potom\ \ M \models A \to C_G B$

předpokládejme

$$M \models A \to E_G(B \wedge A) \quad a \quad (M,s) \models A \qquad (3)$$

indukcí podle k ukážeme, že pro všechny stavy t, dosažitelné z s v k krocích platí

$$(M,t) \models B \wedge A$$

Předpokládejme, že t je G-dosažitelné z s v jednom kroku. Z (3) potom odvodíme

$$(M,s) \models E_G(B \wedge A) \quad \text{odkud} \quad (M,t) \models B \wedge A$$

VZ 2009

Je-li

$$k = k' + 1$$

nechť t' je dosažitelný z s v k' krocích a t v jednom kroku z t'.

Z indukční hypotézy máme

$$(M,t') \models B \wedge A$$

a stejným argumentem jako pro $k = 1$ dostaneme

$$(M,t) \models B \wedge A$$

VZ 2009

Dokázali jsme

$$(M,t) \models B$$

pro každý stav t, G-dosažitelný z s. To znamená

$$(M,s) \models C_G B$$

Ze sémantiky nakonec odvodíme

$$(M,s) \models A \to C_G B$$

Tím je důkaz dokončen.

VZ 2009

Distribuované znalosti

charakterizují znalosti, které odpovídají tomu, když „*všichni agenti dají hlavy dohromady*“. Proto splňují stejné vlastnosti jako znalosti. Navíc upozorníme na dvě drobnosti.

$$\models D_{\{i\}} A \leftrightarrow K_i A$$

pro jednočlennou skupinu se distribuované znalosti shodují s tím, co agent ví.

Naopak, čím je skupina větší, tím větší jsou její distribuované znalosti.

$$Je-li \quad G \subseteq G' \ \ potom \ \ \models D_G A \to D_{G'} A$$

VZ 2009