Formální přístup ke znalostem a jeho aplikace v teorii racionálního jednání

Svatopluk Nevrkla

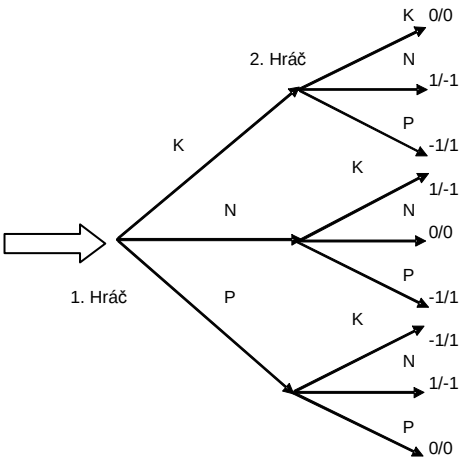
Michal Peliš[*]

Informace a racionální rozhodování

Racionální jednání a rozhodování se často modeluje v rámci teorie her. Přitom hrou se zde nemyslí pouze hry jako šachy či mariáš (i když ty pochopitelně také), ale jakýkoliv dynamický systém, na jehož výstupech se aktivně podílí vícero hráčů/agentů. Hrou tak mohou být různé společenské činnosti, jako vyjednávání o ceně, volba vojenské strategie či vhodné reklamní kampaně. Teorie her má proto mnohé aplikace ve společenských vědách (jde zejména o ekonomii, sociologii a politologii).

Abychom si vysvětlili, co se míní hrou, uvedeme si příklad. Na obrázku 1 je model hry *kámen-nůžky-papír*.

Obrázek 1



Jednotlivé uzly grafu označují možné herní situace, přičemž v kořeni grafu, označeném velkou šipkou, je situace výchozí. U každého uzlu je vyznačeno, který z hráčů je v dané herní situaci na tahu, a z každého uzlu vedou orientované hrany (šipky) označující možné tahy

[*] Text byl zpracován s podporou grantu Dynamické formální systémy (č. IAA900090703, Grantová agentura AV ČR).

tohoto hráče. V listech grafu je pak takzvaná výplatní matice označující, kolik hráči získají bodů.

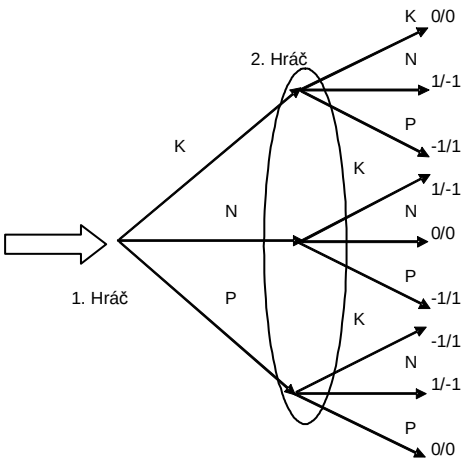
Hra je velmi jednoduchá. Účastní se jí dva hráči a každý z nich má pouze jeden tah. Navíc mají oba hráči v každém svém tahu (nezávisle na herní situaci) vždy stejné možnosti jak táhnout. Vítěz duelu si připíše jeden bod a poražený jeden bod ztrácí. V případě remízy se bodový stav obou hráčů nemění. Jedná se tudíž o (nekooperativní) hru s konstantním (v tomto případě nulovým) součtem. To znamená, že tam, kde jeden hráč body získává, druhý hráč stejný počet bodů ztrácí.

Předmětem studia teorie her pak je, jakou herní strategii má volit racionální hráč za předpokladu, že jeho snahou je vždy maximalizovat bodový zisk. Nás však bude zajímat něco trochu jiného. Povšimněme si, že uvedený model nepopisuje hru *kámen-nůžky-papír* úplně přesně. Ve hře uvedené na nákresu má totiž druhý hráč vždy výherní strategii. To znamená, že jakmile první hráč zvolí libovolný symbol, může druhý hráč vždy zareagovat odpovídajícím symbolem a zvítězit. Takto se ale hra *kámen-nůžky-papír* obvykle nehraje, neboť by byla poměrně nezábavná a pro prvního hráče poněkud frustrující.

Nejběžnější variantou hry je, že oba hráči ukáží symboly naráz, a protože předem nevědí, jaký symbol si vybere jejich spoluhráč, nemohou ani vědět, jak na něj adekvátně zareagovat. Důležité tudíž není, že si hráči vybírají symboly současně, ale že při své volbě předem nevědí o volbě druhého soupeře. Proto by bylo možné tuto hru zrealizovat i tak, že by první hráč nejprve nakreslil zvolený symbol na papír, aniž by jej ukázal soupeři, potom by to samé učinil druhý hráč a následně by své symboly porovnali.

I když první hráč volil svůj symbol jako první, nemělo to na hru žádný vliv, jelikož druhý hráč nevěděl, který ze tří možných symbolů si první hráč zvolil. Nemohl tedy rozlišit od sebe tyto tři různé situace a libovolná volba symbolu by byla stejně odůvodněná jako jakákoliv jiná. Graficky by se tato verze hry znázornila, jak je uvedeno na obrázku 2.

Obrázek 2



Ohraničení kolem tří herních situací 2. hráče naznačuje, že jsou pro něho nerozlišitelné. Pro každého hráče tak mohou být v krajním případě nerozlišitelné všechny situace, kdy je na tahu (pochopitelně za předpokladu, že má v těchto herních situacích stejnou možnost tahů).

Hře, kde jsou všechny herní situace od sebe rozlišitelné, říkáme *hra s úplnou informací*. Takovou hrou jsou například šachy nebo dáma, protože veškeré možné herní situace jsou plně popsány konfiguracemi hracích kamenů na šachovnici, a ty jsou oběma hráčům známé. *Kámen-nůžky-papír* je pak hra s neúplnou informací (dalo by se říci, že se dokonce jedná o hru s nulovou informací, neboť ani jeden z hráčů nedokáže odlišit od sebe žádné dvě situace, kdy je na tahu).

V tomto textu nás však nebudou zajímat hry jako takové. S nimi se lze seznámit např. v [Peliš 2007]. Náš zájem se soustředí jen na *informace* a *znalosti*, které tvoří pozadí každého racionálního jednání a rozhodování.

Jiným příkladem her s neúplnou informací jsou hry karetní. V karetních hrách, jako je například mariáš, určují herní situace distribuce karet. Dobrý hráč mariáše pak dokáže rozlišit daleko více herních situací, jelikož tak nečiní pouze na základě znalostí karet, které vidí ve své ruce, a o kterých již ví, že byly hrány, ale i znalostí pravidel (a předpokladu, že se jimi ostatní hráči řídí) a úvah, jaké znalosti mohou mít ostatní hráči.

Podstatné na úvaze hráče mariáše je, že neuvažuje pouze o všech možných distribucích karet, ale i o tom, co vědí ostatní hráči. Jeho usuzování je proto částečně usuzováním o něčích znalostech. A právě usuzováním o znalostech se zabývá *epistemická logika* (resp. *logika znalostí*), jíž je věnován tento text.

Informace, agenti a situace

Asi nejběžnější je mluvit o znalostech nějakého člověka, ale použití pojmu *znalost* dává smysl i v souvislosti s počítačovými databázemi. Nositelům znalostí se proto v odborné terminologii souhrnně říká *agenti*. Znalosti každého jednotlivého agenta se pak pochopitelně v různých situacích liší, což bude asi nejjednodušší demonstrovat na následujícím příkladě.

Před hráčem skořápek, jenž pro nás bude představovat agenta *i*, leží na stole tři skořápky, přičemž pod každou z nich může (ale nemusí) být ukrytý hrášek. Pokud jediné, co hráče v danou chvíli zajímá, je právě možný výskyt hrášku pod jednou z těchto tří skořápek, bude muset vzít v úvahu osm možných situací. Uvedené situace budeme značit trojicemi:

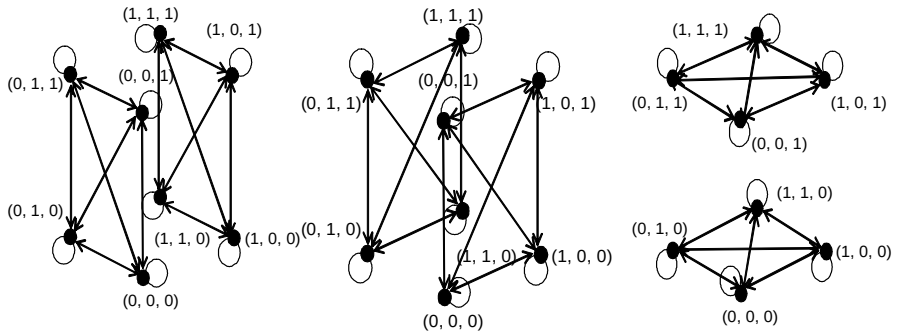
(0,0,0), (1,0,0), (0,1,0), (1,1,0), (0,0,1), (1,0,1), (0,1,1), (1,1,1), kde výskyt jedničky na *n*-té pozici trojice symbolizuje výskyt hrášku pod *n*-tou skořápkou.

Jelikož o skutečném rozložení hrášků nic nevíme, budou pro nás všechny uvedené možnosti představovat možné alternativy. Náš agent však může mít k dispozici více informací (nahlédne-li pod nějakou skořápku) a proto pro něj některé kombinace reálné nebudou. Jelikož nás však zajímají právě *naše znalosti o znalostech tohoto agenta*, bude zapotřebí tuto skutečnost nějak zachytit.

Vyjděme však nejprve z předpokladu, že jsou všechny skořápky neodkryté. Ať už bude skutečné rozložení hrášků pod skořápkami jakékoliv, nebude agent *i* (stejně jako my) schopen uvedené situace od sebe rozlišit. Každá z osmi situací představuje možnou alternativu, a pro agenta jsou nerozlišitelné. Pokud bychom chtěli jeho znalosti popsat pomocí nějakého obrázku, mohli bychom tak učinit například tím způsobem, že bychom každému z možných světů přiřadili nějaký bod a skutečnost, že agent neumí odlišit jeden svět od druhého, bychom vyznačili šipkou od jednoho světa k druhému. V našem případě by bylo možné uspořádat vrcholy do grafu tvaru „kvádru“ a z každého vrcholu by vedla šipka do všech zbývajících sedmi vrcholů (i do tohoto vrcholu samotného). Takový obrázek by ale nebyl příliš přehledný, a proto si raději uvedeme obrázky pro situace, kdy už agent-skořápkář obdržel nějaké informace ohledně možných konfigurací hrášků pod skořápkami.

Na obrázku 3 jsou uvedeny tři různé modely. Tyto modely popisují naše znalosti o znalostech agenta v případě, že o něm víme, že zdvihl levou skořápku (model vlevo), prostřední skořápku (prostřední model), nebo pravou skořápku (model vpravo).

Obrázek 3

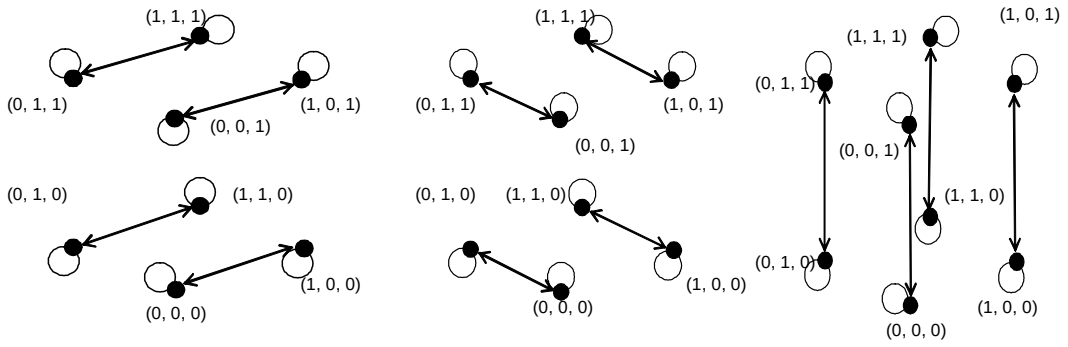


Povšimněme si, že pokud agent *i* zdvihne nějakou skořápku, jeho znalosti se změní v závislosti na skutečnosti, zda pod danou skořápkou byl či nebyl hrášek. Například pokud

zvedne levou skořápku a nalezne pod ní hrášek, pak bude dále pokládat za možné pouze světy (1,0,0), (1,1,0), (1,0,1), (1,1,1), což odpovídá pravé „stěně“ levého modelu. Pokud však hrášek pod levou skořápkou nenalezne, budou pro něj dále připadat v úvahu zbývající světy (0,0,0), (0,1,0), (0,0,1), (0,1,1), odpovídající levé „stěně“ levého modelu. Agent nám (ani čtenáři) neprozradil, zda hrášek pod skořápkou byl (víme pouze, že tato informace je mu dostupná), a proto musíme v modelu nadále zvažovat všech osm světů.

Případ, kdy jsou zdviženy libovolné dvě skořápky je pak na obrázku 4. Na nejlevějším modelu jsou nyní znázorněny znalosti agenta v případě, že jediná skořápka, která nebyla zdvižena, je levá (čili byly zdviženy prostřední a pravá skořápka), na prostředním modelu jsou pak znalosti agenta, pokud zdvihl všechny kromě prostřední skořápky, na pravém modelu jsou znalosti agenta, pokud zdvihl levou a prostřední, nikoliv však pravou, skořápku.

Obrázek 4



Jak vidíme, po zdvihnutí dvou skořápek budou pro samotného agenta zbývat pouze dvě možnosti, jak tomu může být: pod poslední skořápkou je či není hrášek.

V případě, že by byly zdviženy všechny skořápky, agent by měl plnou informaci o distribuci hrášků, a tak by pro něj byl dostupný pouze aktuální svět (to naznačujeme smyčkou u každého z vrcholů), zmizely by už ale všechny zbylé hrany.

Zdvihnutím libovolné skořápky se v každém možném světě příslušného modelu zredukoval počet dostupných možných světů z osmi na čtyři, zdvihnutím druhé skořápky na dva a zdvižením poslední skořápky zůstal dostupný jediný možný svět, a to právě ten aktuální. Čím více má tudíž nějaký agent k dispozici informací, tím lépe dokáže rozlišovat mezi možnými alternativami aktuálního (skutečného) světa. Jinými slovy, světů které nedokáže odlišit od aktuálního je méně.

Modely znalostí

Modely znázorněné na obrázcích 3 a 4 jsou příklady struktury, která se nazývá *kripkovský model* a která je asi nejběžnějším modelem pro popis znalostí.[1] Neformálně můžeme říci, že kripkovský model se skládá z *kripkovského rámce* a *relace splňování*. Kripkovský rámec je definován množinou situací, jimž se obvykle říká *možné světy* (na nákresu reprezentovaných jednotlivými vrcholy) a *relacemi dosažitelnosti* (*nerozlišitelnosti*) na možných světech (reprezentovaných šipkami). V našem příkladě jsme se omezili na jednoho agenta-skořápkáře, a proto jsme hovořili pouze o jedné relaci dosažitelnosti (resp. nerozlišitelnosti). Modelovat však můžeme i stavy znalostí pro více agentů.[2] V takovém případě bude mít každý z agentů svou vlastní relaci dosažitelnosti, což v modelu ošetříme tím, že ke každé šipce napíšeme číslo agenta, do jehož relace dosažitelnosti patří. Pokud je nějaký možný svět *s* v relaci *i* se světem *t* (ze světa *s* vede do světa *t* šipka s číslem *i*), řekneme že *t* je *i-dosažitelný* z *s*.

Relace splňování je pak přiřazením výroků (faktů) k možným světům (situacím). Ve výše uvedeném příkladě relace splňování určuje příslušnost jednotlivých „konfigurací hrášků“ ke všem uzlům grafu. K formálním definicím uvedených pojmů se ještě dostaneme.

Čtenář již možná zpozoroval, že přestože jsme si uvedli několik různých kripkovských modelů, měla relace dosažitelnosti ve všech uvedených rámcích shodné vlastnosti. Ty vyplývají z toho, že relací dosažitelnosti chceme popsat skutečnost, že nějaké dva světy jsou pro daného agenta nerozlišitelné. Zamysleme se nyní nad pojmem *nerozlišitelnosti*.

Jelikož je každý svět nerozlišitelný sám od sebe, musí být sám ze sebe i dosažitelný. Libovolná relace dosažitelnosti tudíž bude vždy *reflexivní*, čili z libovolného světa “povede vždy kruhová šipka“ do něho samotného.

Pokud je jeden svět nerozlišitelný od druhého, je zákonitě i druhý nerozlišitelný od prvního. Z tohoto důvodu je jasné, že je-li ze světa *s* dosažitelný svět *t*, pak i z *t* je dosažitelný svět *s* (všechny šipky nákresu jsou oboustranné, kromě kruhových šipek, kde by se jednalo o pouhé zdvojení). Matematickou terminologií řekneme, že relace dosažitelnosti je *symetrická*.

Dále si představme tři situace takové, že první je neodlišitelná od druhé a druhá od třetí. Pak zákonitě musí být i třetí situace nerozlišitelná od první. Čili vede-li šipka ze světa

[1] Název je odvozen od jména známého logika a filosofa Saula Kripka (nar. 1940).
[2] Epistemické logiky (podobně jako teorie her) obvykle pracují s konečným počtem agentů.

s do světa *t*, a ze světa *t* vede šipka do světa *u*, pak vede šipka i ze světa *s* do světa *u*. Relaci splňující uvedenou vlastnost nazveme *tranzitivní*.

Spojením všech tří vlastností (*reflexivita*, *symetrie*, *tranzitivita*) získáme relaci *ekvivalence*. Ekvivalence tedy splňuje následující:

- *Reflexivita* Pro libovolný svět *s* platí: *s* je dosažitelný z *s*.
- *Symetrie* Pro libovolné dva světy *s,t* platí: Je-li *s* dosažitelný z *t*, pak je i *t* dosažitelný z *s*.
- *Tranzitivita* Pro libovolné světy *s,t,u* platí: Je-li *s* dosažitelný z *t* a zároveň je *t* dosažitelný z *u*, pak je i *s* dosažitelný z *u*.

Bude-li relace dosažitelnosti ekvivalencí, pak bude odpovídajícím způsobem zachycovat, že některé možné světy jsou od sebe nerozlišitelné.[3]

Kripkovským rámcem pro epistemickou logiku budeme rozumět takovou strukturu, jejíž všechny relace dosažitelnosti jsou ekvivalencemi.[4] Tím dospíváme k následující definici:

Definice 1 (kripkovský rámec)

Kripkovský rámec **R** s *n* relacemi dosažitelnosti je uspořádanou *n+1*-ticí $<S, \sim_1,\ldots,\sim_n>$, kde *S* je (libovolná) množina možných světů a $\sim_1 \ldots \sim_n$ libovolné ekvivalence na *S*.
Platí-li $s \sim_i t$ pro nějaké dva světy *s*, *t* z *S*, pak řekneme, že *t* je *i-dosažitelný* z *s*.

Uveďme si nyní ještě příklad kripkovského modelu, ve kterém modelujeme znalosti více agentů. Představme si, že na stole leží rubem vzhůru tři karty. Na lících karet jsou zapsána tři písmena. Na první je A, na druhé B a na třetí C. Dva hráči, kteří nyní představují agenty 1 a 2, si vyberou po jedné kartě a podívají se na písmeno, které je na ní napsáno, aniž by věděli, jakou kartu si vytáhl druhý hráč a která zůstala ležet na stole.

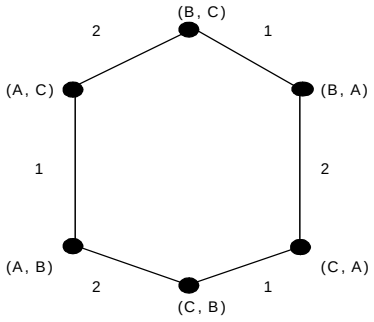
My samozřejmě nevíme ani jakou kartu si který z obou hráčů vytáhl, víme však jaké jsou možné situace. Možných světů je nyní pouze šest: (A,B), (A,C), (B,A), (B,C), (C,A), (C,B). V každém světě pak má každý z agentů k dispozici jinou informaci a my můžeme uvažovat o tom, co v nich tito agenti vědí.

[3] Existuje jednoznačná korespondence mezi ekvivalencemi na nějaké množině a rozklady dané množiny. Třídy rozkladů množiny možných světů v rozkladu korespondujícím s relací dosažitelnosti nějakého agenta pak jsou množinami světů, které jsou pro daného agenta navzájem nerozlišitelné. Další motivace volby relace dosažitelnosti jako ekvivalence bude naznačena při diskusi epistemických formulí platných v každém kripkovském modelu pro epistemickou logiku.

[4] Na dalších obrázcích budeme automaticky vynechávat kruhové šipky plynoucí z reflexivity a místo šipek vedoucích oběma směry (symetrie) budeme kreslit pouze plnou čáru.

Tak například ve světě (A,B), který teď považujme za skutečný, vidí agent 1, že v rukách drží kartu A, ale neví, zda má agent 2 v rukou kartu B či C. Pro agenta 1 ve světě (A,B) proto připadají v úvahu možné světy (A,B) a (A,C). Oproti tomu agent 2 ví, že má kartu B, ale zase neví, zda si agent 1 vytáhl kartu A, nebo kartu C. Proto pokládá za možné dva světy a to (A,B) a (C,B). Jinými slovy, světy (A,B) a (A,C) jsou ze světa (A,B) 1-dosažitelné a světy (A,B) a (C,B) jsou ze světa (A,B) 2-dosažitelné. Úplný popis znalostí obou agentů je zachycen na obrázku 5.

Obrázek 5



Formalizace výroků o znalostech a kripkovské modely

V minulé kapitole jsme si představili různé kripkovské rámce, kde jsme každému možnému světu intuitivně přiřadili nějakou situaci pomocí *relace splňování*. V jednom případě jsme tak možným světům přiřadili možné distribuce hrášků pod třemi kalíšky, ve druhém distribuce tří karet mezi dvěma hráči. Abychom však dospěli ke zcela obecnému popisu *kripkovského modelu*, budeme si muset zvolit nějaký pevný elementární jazyk, který bude vhodný k popisu jak všech možných faktických situací, tak i znalostí agentů. K efektivnějšímu popisu znalostí agentů si proto definujme formální jazyk epistemických logik.

Pro zjednodušení budeme předpokládat, že všechny relevantní informace, které mohou nějak ovlivnit usuzování a jednání agentů, jsou popsatelné prostředky výrokové logiky. Budeme proto pracovat s množinou výrokových atomů, které budeme obvykle značit malými písmeny latinské abecedy (*p*, *q*, *r*,...). Výrokové atomy budou označovat (navzájem různé) primitivní výroky ohledně faktického světa.[5]

[5] Jaké výroky přirozeného jazyka si zvolíme jako primitivní je na nás.

Výrokové atomy pak můžeme kombinovat pomocí výrokových spojek $\neg$, $\wedge$ ve výrokové formule, jako např. $\neg p$, $p \wedge \neg q$, $\neg(p \wedge p)$, které budeme značit velkými písmeny latinské abecedy (*A, B, C,...*). Formální definice výrokových formulí je následující:

Definice 2 (výroková formule)

- Každý výrokový atom je výrokovou formulí.
- Je-li *A* výrokovou formulí, je i $\neg A$ výrokovou formulí.
- Jsou-li *A* a *B* výrokové formule, pak i $(A \wedge B)$ je výrokovou formulí.[6]

Epistemickou výrokovou formuli pak získáme pomocí epistemického operátoru K_i před formulí *A*.[7] Formuli K_iA budeme číst jako

agent *i* ví, že platí *A*.

Definice 3 (epistemická výroková formule)

- Každá výroková formule (Definice 2) je epistemická výroková formule.
- Je-li *A* epistemická výroková formule a *i* přirozené číslo, je K_iA epistemická výroková formule.

V předchozí kapitole jsme chápali možné světy jakožto vybrané situace reálného světa. Uvedené situace můžeme nyní identifikovat s množinami atomických výroků, které „za dané situace *platí*".

Označme výrok *Pod první skořápkou je hrášek.* symbolem H_1, výrok *Pod druhou skořápkou je hrášek.* symbolem H_2 a výrok *Pod třetí skořápkou je hrášek.* symbolem H_3. Například svět (0,1,1) ve „skořápkovém modelu" identifikujeme s množinou atomických výroků $\{H_2, H_3\}$, svět (0,0,0) s prázdnou množinou a svět (0,0,1) s jednoprvkovou množinou $\{H_3\}$.

Podobně můžeme situaci na obrázku 5 popsat šesti výrokovými atomy A_1, B_1, C_1, A_2, B_2, C_2, kde například atom B_1 znamená *První hráč má v ruce kartu B.* a atom A_2 *Druhý hráč má v ruce kartu A.*

[6] Pro větší přehlednost notace budeme vynechávat vnější závorky a zavedeme následující zkratky: $A \vee B$ pro $\neg(\neg A \wedge \neg B)$ a $A \rightarrow B$ pro $\neg(A \wedge \neg B)$.
[7] *K* je z anglického slova *knowledge*.

Splňování výrokových formulí bude definováno v souladu se sémantikou spojek podle pravidel pro klasickou výrokovou logiku (body 1) a 2) v následující definici). Důvod proč vůbec zavádět kripkovské rámce pro epistemickou logiku vyplyne z bodu 3) téže definice.

Definice 4 (kripkovský model)

Kripkovský model **M** definujeme jako dvojici $<\Vdash, \mathbf{R}>$, kde **R** je kripkovský rámec a $\Vdash$ relace splňování mezi světem *s* (daného rámce) a epistemickými výrokovými formulemi.

Je-li *A* epistemická výroková formule, zápis $s \Vdash A$ budeme číst: „Formule *A platí* (je *splněna*, resp. je *pravdivá*) ve světě *s* (modelu **M**)“. Pro uvedenou relaci splňování bude navíc platit:

1) Pro libovolné *s*: $s \Vdash A$ právě tehdy, když není pravda, že $s \Vdash \neg A$
2) Pro libovolné *s*: $s \Vdash A \wedge B$ právě tehdy, když $s \Vdash A$ a zároveň $s \Vdash B$
3) Pro libovolné *s*: $s \Vdash K_i A$ právě tehdy, když pro libovolný svět *t* i-dosažitelný z *s* (v daném rámci) platí $t \Vdash A$

Dohodněme se nyní, že kripkovským modelům budeme občas říkat modely a kripkovským rámcům budeme příležitostně říkat pouze rámce. Řekneme, že nějaký model **M** je *modelem nad rámcem* **R**, jestliže **M** je $<\Vdash, \mathbf{R}>$.

Z uvedené definice je jasné, že chceme-li popsat relaci splňování v některém světě, stačí se omezit na výpis výrokových atomů, které jsou v daném světě splněny. Z vlastností relace splňování nám pak zcela jednoznačně vyplyne, zda budou v daném světě splněny všechny zbývající formule.

Vhodná volba definic dalších výrokových spojek pomocí spojek negace a konjunkce nám pak zaručí, že relace splňování bude mít i následující vlastnosti:

4) Pro libovolné *s* platí: $s \Vdash A \vee B$ právě tehdy, když $s \Vdash A$ anebo $s \Vdash B$
5) Pro libovolné *s* platí: $s \Vdash A \rightarrow B$ právě tehdy, když jestliže $s \Vdash A$, pak i $s \Vdash B$

Zastavme se nyní u bodu 3), který definuje, kdy je v nějakém světě splněna epistemická výroková formule a objasňuje důvod zavádění kripkovských rámců. Platnost všech epistemických výrokových formulí je vždy odvozena i z odpovídajících relací dosažitelnosti. Intuitivní motivace bodu 3) je, že agent *i* ví, že platí výrok *A* právě tehdy, když *A* platí ve všech světech, které *i* pokládá za možné, čili nedokáže si na základě svých

informací představit jako možný nějaký svět *s*, ve kterém by *A* neplatilo. To nastane například v případě, kdy *i* nahlédne pod prostřední skořápku a zahlédne pod ní hrášek (není-li slepý, nepozorný, či výjimečně nedovtipný), jelikož nyní už žádný svět, ve kterém by nebyl pod prostřední skořápkou hrášek nepokládá za možný. Jinými slovy ve všech světech, které *i* pokládá za možné, je pod druhou skořápkou hrášek. Čili, podle naší definice, *i* ví, že pod prostřední skořápkou je hrášek.

Při konstrukci modelu na obrázku 5 jsme vycházeli z jistých předpokladů, které musí daný model splňovat. Například jsme implicitně vycházeli z faktu, že v každém možném světě má první hráč v ruce vždy jednu kartu ze tří možných, čili v každém světě platila disjunkce $A_1 \vee B_1 \vee C_1$. Také druhý hráč měl vždy jednu z těchto karet. V každém světě platilo i $A_2 \vee B_2 \vee C_2$. Platí-li nějaká formule v libovolném světě daného modelu, řekneme prostě, že platí v celém modelu:

Definice 5

Řekneme, že formule A *platí* v modelu **M** (je *pravdivá* v modelu **M**), pokud platí v každém světě modelu **M**:

$$\mathbf{M} \models A \text{ právě tehdy, když pro každé } s \in S \text{ platí: } s \Vdash A$$

Řekneme, že formule A je *splnitelná* v modelu **M**, pokud platí alespoň v jednom světě modelu **M**. V opačném případě je *nesplnitelná* (tj. neplatí v žádném světě modelu **M**).

Uvědomme si nyní, že jsme doposud při konstrukci modelů požadovali, aby se možné světy lišily pouze určitými výrokovými atomy. Tak například na obrázku 3 jsme rozlišovali různé situace pouze podle možné distribuce hrášků pod skořápkami, čili podle platnosti atomů H_1, H_2, H_3. Nezajímalo nás například, v jakém z možných světů například prší. Při konstrukci modelu je proto vhodné si nejprve ujasnit, které výrokové atomy jsou relevantní pro reprezentaci problému, jaké jsou jejich přípustné kombinace splňující zadání úlohy (jaké jsou možné světy) a potom určit relaci dosažitelnosti pro každého agenta.

Jelikož chceme rozlišovat možné světy pouze na základě vybraných výroků, bude zapotřebí libovolný jiný výrok splnit buď ve všech světech modelu nebo v žádném. Později uvidíme, že výroky platné ve všech světech jsou obecnou znalostí. Zajímavější je ale zamýšlet se nad tím, které výroky musíme splnit v každém světě, abychom dostali věrnou reprezentaci znalostí.

Představme si nyní rámec, který má libovolný počet různých světů, ale z každého světa *s* je dosažitelný pouze on sám (i taková relace dosažitelnosti je ekvivalencí). Uvažujme nyní platnost výrokového atomu *p* v některém ze světů *s* daného rámce. Jsou pouze dvě možnosti: buď v *s* bude *p* platit, anebo nikoliv. Jelikož je z každého světa *s* rámce **R** dosažitelný pouze sám svět *s*, bude v *s* platit $K_1(p)$, resp. $K_1(\neg p)$. Pro libovolnou volbu relace splňování pak bude v libovolném světě *s* rámce **R** platit disjunkce $K_1(p) \vee K_1(\neg p)$. Podle definice proto bude formule $K_1(p) \vee K_1(\neg p)$ platit v celém modelu, ať už relaci splňování zvolíme jakkoliv. Proto má smysl říci, že formule $K_1(p) \vee K_1(\neg p)$ platí v rámci **R**.

Definice 6

Řekneme, že formule *A platí* (je *pravdivá*) v rámci **R**, pokud platí v každém modelu nad rámcem **R**:

$\mathbf{R} \models A$ právě tehdy, když pro libovolný model **M** nad rámcem **R** platí: $\mathbf{M} \models A$

V rámcích, kde je nějaká relace dosažitelnosti pro agenta *i* identitou, čili vede z každého světa pouze do něho samotného, je agent *i* vševědoucí. V takových rámcích pro libovolnou formuli *A* platí $A \rightarrow K_i(A)$, čili pokud nějaká skutečnost nastává, agent *i* o ní ví. V rámcích s „bohatší“ relací dosažitelnosti však již formule $K_1(p) \vee K_1(\neg p)$ platit nemusí. Přesto však v daném rámci platí některé jiné formule, jejichž platnost je zaručena vlastnostmi relace dosažitelnosti.

Ve všech rámcích, které dosud uvažujeme, platí například každá formule tvaru

$$K_i A \rightarrow A$$

pro libovolně zvoleného agenta *i* a libovolnou formuli *A* (čili např. formule $K_1 p \rightarrow p$, $K_2 K_1 p \rightarrow K_1 p$, $K_2(p \rightarrow q) \rightarrow (p \rightarrow q)$ apod.). Výše uvedené schéma se obvykle nazývá *axiom* **T**.

Zamyslíme-li se nad důvodem platnosti uvedeného schématu, zjistíme, že platí, jelikož relace dosažitelnosti je reflexivní. Pokud v nějakém světě *s* platí $K_i A$, platí podle definice *A* ve všech světech, jež jsou i-dosažitelné ze světa *s*. Protože však zmíněná relace je reflexivní je i svět *s* i-dosažitelný sám ze sebe. Proto každá formule, která vznikne dosazením libovolné formule za *A* do schématu $K_i A \rightarrow A$, platí v libovolném epistemickém rámci.[8] Neformální význam schématu **T** pak je:

Jakékoliv poznání (faktu) jakéhokoliv (racionálního) agenta je vždy pravdivé.

[8] Výsledkům zmíněné substituce potom říkáme *instance axiomu* **T**.

V rámcích pro *doxastickou logiku* (logiku přesvědčení) pak samozřejmě schéma **T** obecně platit nemůže, jelikož existují i nepravdivá přesvědčení. Obvykle se proto v doxastických logikách nahrazuje axiom **T** slabším schématem **D**, které zaručuje, že přesvědčení racionálního agenta jsou konzistentní a má následující formu:

$$K_i A \rightarrow \neg K_i \neg A$$ [9]

Čtenáře nyní jistě napadne, jestli další vlastnosti relací dosažitelnosti v epistemických rámcích nezaručí platnost nějakých dalších axiomových schémat. Tranzitivita relace dostupnosti zaručí platnost schématu **4**, které má tvar:

$$K_i A \rightarrow K_i K_i A$$

a kterému se říká axiom *pozitivní introspekce*; agent si je vědom svých znalostí:

Když (racionální) agent ví A, pak též ví, že ví A.

Dalším schématem, které platí v každém rámci, jehož relace je symetrická, je schéma **B**:

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

Schéma **B** bývá v reflexivních rámcích nahrazováno poněkud přehlednějším schématem *negativní introspekce*, které se značí **5** a má formu

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

Platnost *negativní introspekce* zaručuje, že agentovi je známa jeho neznalost.

Existují však i formule, které platí v každém rámci (bez ohledu na relaci dosažitelnosti). Univerzálně platným axiomem je například schéma **K**:

$$K_i(A \rightarrow B) \rightarrow (K_i A \rightarrow K_i B),$$

které vyjadřuje epistemickou variantu pravidla známého v logice pod jménem *modus ponens*; ví-li agent, že *jestliže A, pak B*, pak, když se dozví, že *A*, může usoudit, že i *B*. Všechny instance schémat **K**, **T**, **4**, **B, 5** jsou proto příklady formulí, které platí v každém rámci pro námi zavedenou epistemickou logiku.

Skupinové znalosti

Doposud jsme se zabývali pouze znalostmi jednotlivých agentů jakožto nezávislých individuí. Podstatnou část předmětu epistemických logik však tvoří zkoumání znalostí, které jsou sdílené, obecně přijímané nebo distribuované v určité skupině agentů. Před zavedením

[9] Pro zajímavost: Dá se ukázat, že pokud přeložíme výrok "Agent *i* ví A", jako "Agent *i* je přesvědčen, že A a zároveň A je pravda." pak obě logiky dokazují „stejné" výroky.

formálních definic, bude vhodné uvést si pár příkladů z běžného života, ve kterém se tyto pojmy uplatňují.

Asi nejintuitivnějším pojmem je *sdílená znalost*. Znalost sdílená v nějaké skupině agentů je taková znalost, kterou mají všichni agenti této skupiny. Pokud si například tento text přečtou všichni studenti sociologie na Filozofické fakultě, bude jeho obsah sdílenou znalostí. Mohlo by se zdát, že sdílená znalost je dostatečným předpokladem pro úspěšnou komunikaci či kooperativní jednání. To však není tak docela pravda.

Představme si, že by agenti v nějaké skupině, řekněme všichni občané České Republiky, sdíleli znalost, že na přechodu se má na červenou zastavit, což by věděli jak chodci, tak řidiči. Z toho by však nijak neplynulo, že by všichni občané České Republiky také sdíleli znalost o tom, že toto pravidlo silničního provozu je sdílenou znalostí. Chodec na přechodu by tak sice viděl, že na semaforu pro automobily svítí červená a že by automobily měly zastavit, ale nevěděl by jistě, jestli to také ví všichni řidiči. Kdyby se tedy k přechodu blížilo auto, nevěděl by chodec, zda řidič zastaví. Řidič možná červenou na semaforu zpozoroval, ale nechápe ji jako signál k tomu, že má nechat chodce přejít. Dejme tedy tomu, že by chodec věděl, že řidič ví, že má zastavit. Nyní by však řidič nevěděl, že to chodec ví a nevěděl by, zda se odváží na přechod vkročit. Rozumný řidič by proto asi přeci jen zastavil a nechal chodce přejít, ať už by věděl, že je to kvůli červené anebo nikoliv. Jestliže by pak na semaforu svítila pro řidiče zelená, nevěděl by, zda i chodci ví, že nyní na přechod nesmí a musel by na přechodu zastavovat vždy. Kdyby pak proti sobě nestáli automobilista a chodec, ale například dvě cisterny s výbušninou, asi by se musela křižovatka zablokovat. Sdílená znalost proto není vždy dostatečným předpokladem úspěšné komunikace a koordinace jednání, neboť někdy je zapotřebí také sdílená znalost, že něco je sdílenou znalostí, nebo dokonce sdílená znalost, že je sdílenou znalostí, že něco je sdílenou znalostí ... (apod.) Takové znalosti, které splňují všechny požadavky tohoto druhu se nazývají *obecná znalost*.

Velmi známým příkladem problému obecné znalosti je *problém koordinovaného útoku*:

> Dva generálové se svými vojsky táboří na dvou kopcích nad údolím, ve kterém je nepřítel. Oba sdílejí obecnou znalost, že musí zaútočit v přibližně stejnou dobu, aby nepřítele porazili, a že pokud zaútočí pouze jeden z nich, nepřítel jeho vojsko rozdrtí. Z jednoho kopce však na druhý není vidět, a proto jediným možným způsobem komunikace je posílání depeší po poslovi. Jelikož však mezi tábory obou generálů leží tábor nepřítele, existuje jisté riziko, že nepřítel posla s depeší zadrží. První z generálů pošle depeši s přesnou hodinou útoku druhému generálovi. Jelikož však neví, zda

depeše dorazila, nebude útočit, pokud si nebude jistý, že druhý generál depeši obdržel a že útok skutečně provede v určenou hodinu. Druhý generál může sice po obdržení zprávy poslat posla nazpět se zprávou, že původní vzkaz dorazil. Pak však zase nebude vědět, zda první generál skutečně dostal jeho ujištění a zda nebude s útokem váhat, jelikož neví, zda druhý generál ví o plánovaném útoku.

Takto si mohou tito generálové posílat poslů kolik chtějí, ale nikdy nedosáhnou obecné znalosti o hodině útoku a jejich útok bude vždy rizikem.

Obecné znalosti však hrají důležitou roli také v teorii her. Během hraní her totiž hráči při svém usuzování často vychází z předpokladu, že všichni účastníci hry sdílejí extenzivní obecné znalosti. Takovými znalostmi nejsou pouze znalosti pravidel hry, ale například znalost jejího dosavadního průběhu.

Dalším zajímavým pojmem je *distribuovaná znalost*. Znalosti distribuované mezi nějakou skupinou agentů jsou pak takové znalosti, jaké by měl nějaký agent, kdyby měl k dispozici všechny informace, co mají agenti této skupiny. Tak například v karetní hře je mezi jejími hráči distribuována znalost rozložení karet v rukách hráčů. Každý jednotlivý hráč ví pouze o kartách ve své ruce. Kdyby však nějaký kibic nahlížel do rukou všech hráčů, měl by k dispozici informace, které v plném rozsahu nemá k dispozici žádný hráč, ale které jsou mezi těmito hráči distribuované.

Jedním z fenoménů současné společnosti je, že s narůstajícím obsahem vědění rostou i požadavky na úzkou specializaci vědců a odborníků. Renesanční představa člověka-všeuměla je již dávno nerealizovatelná. Roste proto úloha vědeckých týmů a komunikace mezi členy tohoto týmu, neboť tak je možno šířit distribuované znalosti. Mezi různými vědci-specialisty potom mohou být distribuované veškeré podstatné informace k řešení nějakého konkrétního problému.

Následující hádanka *Tři mudrci* nám poslouží jako další motivace k zavedení pojmu obecné znalosti:

Nudící se vladař povolal své tři nejlepší mudrce, ukázal jim tři černé a tři bílé klobouky a oznámil jim:

“Nyní vám všem zaváži oči a nasadím vám na hlavu každému jeden z těchto klobouků. Zbylé klobouky pak schovám a všem vám oči zase rozváži, abyste všichni viděli, jaké klobouky mají zbývající dva mudrci na svých hlavách, ale nemohli usoudit, který klobouk zbyl na vás.“

Jak král řekl, tak také učinil. Když rozvázal všem mudrcům oči, prohlásil:

“Nejméně jeden z vás má na hlavě černý klobouk. Kdo si myslíte, že máte na hlavě černý klobouk, zdvihněte ruku!“

Žádný z mudrců se neozval. Král proto řekl:

“Říkám vám znovu: Kdo si myslíte, že máte na hlavě černý klobouk, zdvihněte ruku!“

Mudrci opět nehybně stáli a hleděli na krále. Král tedy znovu promluvil:

“Říkám vám do třetice: Kdo si myslíte, že máte černý klobouk, zdvihněte ruku!“

Tentokrát však všichni tři mudrci zdvihli ruku a skutečně všichni měli na hlavě černé klobouky.

Jak mohli mudrci dospět k takovému závěru? A jak je možné, že k němu dospěli, až když se jich král zeptal potřetí? V následující úvaze si ukážeme, že je-li n počet mudrců s černým kloboukem na hlavě, pak na prvních $n–1$ královych otázek nezareaguje nikdo a na další otázku pak správně zareagují už všichni mudrci s černým kloboukem.

Předpokládejme, že by černý klobouk měl na hlavě pouze jeden mudrc. Pak by na hlavách zbylých dvou mudrců viděl bílé klobouky. Z králova oznámení, že alespoň jeden z nich má na hlavě černý klobouk, by proto okamžitě usoudil, že to musí být on.

V případě, že by král rozmístil na hlavy mudrců dva černé klobouky, tak by na svou první otázku nedostal odpověď, protože všichni mudrci by viděli alespoň jeden černý klobouk na hlavě jiného mudrce. Při druhé králově otázce by však mudrci s černými klobouky na hlavě mohli usuzovat takto:

„Vidím jeden černý a jeden bílý klobouk na hlavách svých kolegů. Když se však král poprvé ptal, zda někdo neví o černém klobouku na své hlavě, mudrc s černým kloboukem na hlavě mlčel. To znamená, že musel vidět alespoň ještě jeden černý klobouk, jinak by musel usoudit, že král mluví o něm. Jelikož však druhý z mudrců má bílý klobouk, musel vidět černý klobouk na mé hlavě.“

Oba mudrcové s černými klobouky na hlavě by se proto přihlásili.

V případě, že by král rozdal všem mudrcům černé klobouky, bylo by jejich usuzování po třetí otázce obdobné. Při druhé otázce by si všichni mohli říci:

„Oba zbylí mudrci mlčeli při první otázce, protože viděli černý klobouk na hlavách toho druhého, ale když se král zeptal podruhé a oni mlčeli, nemohu mít na hlavě bílý klobouk, jelikož pak by se oba přihlásili. Musím tedy mít černý klobouk na hlavě i já.“

Podobnou úvahu lze užít pro libovolný počet mudrců.

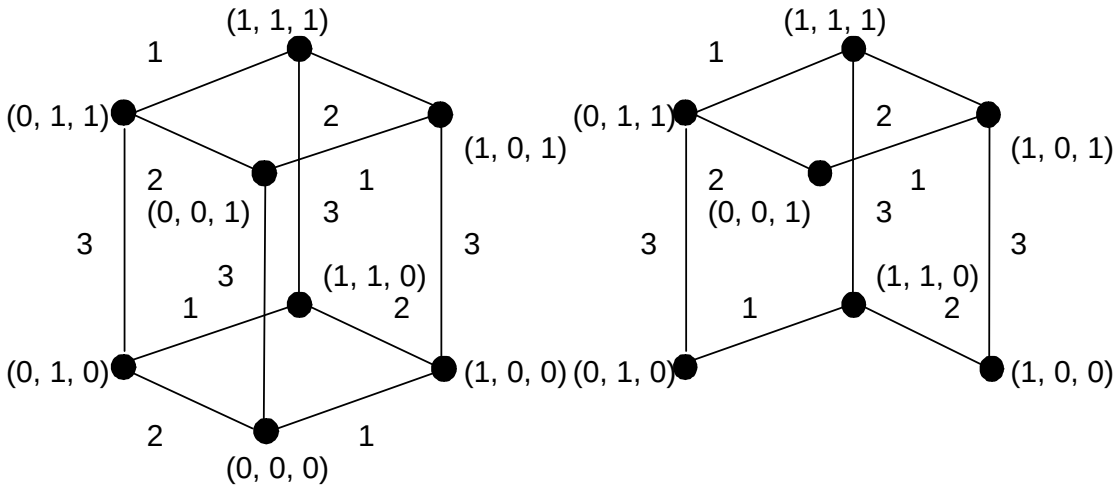
Zamysleme se však nad tím, co by mohli mudrci usoudit, kdyby jim vladař ihned na počátku neprozradil, že alespoň jeden z nich má na hlavě černý klobouk. Po první králově otázce by žádný z mudrců nevěděl, zda má na hlavě černý klobouk, ani v případě, že by

kolem sebe viděl jenom mudrce s bílými klobouky (všichni ví, že král klidně mohl rozdat všem bílé klobouky). Všichni mudrci by navíc věděli, že nikdo z nich neví, zda má na hlavě černý klobouk, či ne. Kdyby se proto král zeptal znovu (a znovu a znovu, libovolněkrát…) a žádný z mudrců se nehlásil, pak by všichni mudrci věděli, že je to proto, že nikdo nemůže vědět o barvě klobouků na své hlavě, ať je rozložení klobouků jakékoliv. Proto by ani po libovolném počtu královych otázek nemohl žádný z mudrců definitivně usoudit, zda má na hlavě černý klobouk či nikoliv.

To však v důsledku znamená, že královo oznámení je nějakým způsobem informativní, ačkoliv v případě, že alespoň dva mudrci mají černé klobouky, vlastně král říká něco, co všichni vědí. V případě, že alespoň tři mudrci mají černé klobouky, říká král dokonce něco, o čem všichni vědí, že to všichni vědí. Královo oznámení však bude informativní, i v případě, že bude mudrců libovolný počet a libovolný počet z nich bude mít na hlavě černé klobouky.

Jak je ale možné, že královo jednoduché oznámení o přítomnosti jediného černého klobouku vždy obsahuje více informací, než kolik může libovolný mudrc obdržet shlédnutím libovolně velkého počtu černých klobouků na hlavách ostatních mudrců? Na obrázku 6 jsou modely znalostí mudrců před a po králově oznámení.

Obrázek 6



Na základě znalostí z předchozí kapitoly můžeme pozorovat, že v obou modelech platí určité předpoklady. Všichni mudrci samozřejmě vědí, jaké jsou možné kombinace klobouků na

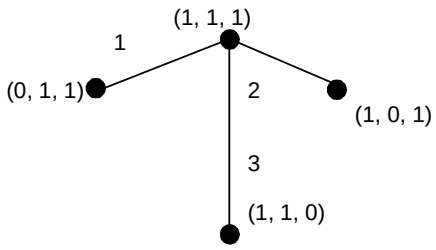
jejich hlavách, ale žádný z nich nevidí na svou hlavu a neví, jaký klobouk má. Zároveň je patrné, že ve všech světech, kde král rozmístil na hlavy mudrců alespoň dva černé klobouky, tj. (0,1,1), (1,0,1), (0,1,1) a (1,1,1), každý z mudrců ví, že alespoň jeden z mudrců má na hlavě jeden černý klobouk, a proto (0,0,0) není z žádného z těchto světů dostupný ani pro jednoho z mudrců.

Předpokládejme nyní, že král oznámí, že alespoň jeden z mudrců má na hlavě černý klobouk (všichni mudrci oznámení slyší, rozumí mu a předpokládají, že oznámení je pravdivé). Tato situace je znázorněna v pravé části obrázku 6. Svět (0,0,0) nemůže v modelu zůstat, protože by byl dostupný sám ze sebe. Pokud však předpokládáme, že král nelhal a že jeho oznámení je pravdivé a informativní, pak svět (0,0,0), ve kterém by všichni mudrci věděli, že nikdo z nich nemá na hlavě černý klobouk, není přípustnou možností.

Na uvedeném modelu je patrné, že po králově oznámení v situaci, kdy má na hlavě černý klobouk pouze jeden mudrc, všichni (tudíž i on), vědí, jaké je rozložení kloboukú. Avšak nové informace mají všichni i v ostatních světech. Už před královým ohlášením věděli mudrci ve světech (0,1,1), (1,0,1) a (0,1,1), že alespoň jeden mudrc má na hlavě černý klobouk. Po králově oznámení však všichni dokonce ví, že všichni ví, že alespoň jeden mudrc má na hlavě černý klobouk, což předtím nevěděli, protože se mohli domnívat, že mudrc s černým kloboukem na hlavě vidí pouze bílé klobouky a neví o existenci černého klobouku na své hlavě. Tuto znalost měli před královým ohlášením všichni mudrci pouze ve světě (1,1,1), neboť všichni věděli, že každý z mudrců vidí černý klobouk na hlavě alespoň jednoho dalšího mudrce. Nevěděli však, zda každý vidí takové klobouky dva, protože si nebyli jisti barvou klobouku na své hlavě. Z toho důvodu nemohli vědět, že všichni vědí, že všichni vědí, že jeden z mudrců má na hlavě černý klobouk. Královo oznámení je proto informativní pro každého mudrce, ať už kolem sebe vidí libovolný počet černých kloboukú. Veřejným královým ohlášením faktu, že alespoň jeden z mudrců má černý klobouk, se tento fakt stává obecnou znalostí všech mudrců, což je podstatně více než pouhá společná znalost tohoto faktu.

Na obrázku 7 jsou znázorněny znalosti mudrců po první králově výzvě.

Obrázek 7

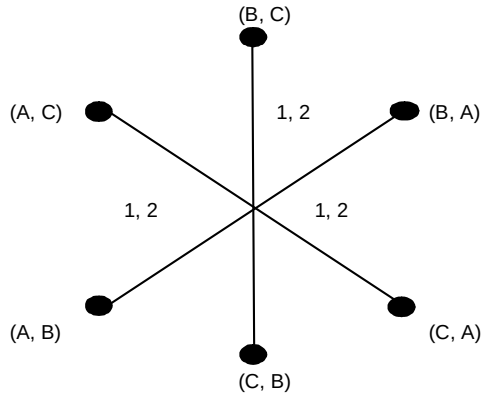


Jakmile první králova výzva zůstane bez odpovědi, všichni mudrci vědí, že žádný z nich nemůže vědět, že má na hlavě černý klobouk. To by však nebyla pravda ve světech, kde by byl pouze jeden mudrc s černým kloboukem na hlavě, tj. (1,0,0), (0,1,0) a (0,0,1). Také ony přestanou být věrohodnou alternativou skutečnosti. Při další králově výzvě pak zůstane v modelu pouze jediný aktuální a sám ze sebe dostupný svět (1,1,1).

Formální vlastnosti skupinových znalostí

Představme si však nyní alternativu situace zachycené modelem na obrázku 5. Oba agenti si na začátku vyberou jednu kartu, ale tentokrát se na ni ani nepodívají. Místo toho se otočí třetí karta. Uvedená situace je znázorněna modelem na obrázku 8.

Obrázek 8



Z uvedeného modelu je jasné, že nyní se znalosti obou hráčů zcela shodují, neboť mají úplně stejné informace. V každém možném světě pokládají oba agenti za možné světy s libovolnou distribucí zbývajících dvou karet neboť oba vědí o kartě, kterou nedrží žádný z nich. Takže ve světech (A,B), ale i (B,A) platí $K_1(\neg C_1 \wedge \neg C_2)$ i $K_2(\neg C_1 \wedge \neg C_2)$. V těchto světech je proto skutečnost $\neg C_1 \wedge \neg C_2$ *sdílenou znalostí*. Definice sdílené znalosti je následující:

Definice 7 (sdílená znalost)

Nechť G je podmnožina množiny agentů. Zaveďme nyní do jazyka epistemických logik nový modální operátor E_G pro *sdílenou znalost.*[10]

Pokud G bude množina všech agentů, budeme místo $E_G A$ psát prostě EA.

Splňování definujeme následovně:

6) Pro libovolné *s* platí: $s \Vdash E_G A$ právě tehdy, když $s \Vdash K_i A$ pro každého agenta $i \in G$

Výrok $\neg C_1 \wedge \neg C_2$ však není jen sdílenou znalostí, ale dokonce *znalostí obecnou*.

Definice 8 (obecná znalost)

Buď G podmnožina množiny agentů. Obohaťme jazyk epistemické logiky o modální operátor C_G.[11]

Je-li G množina všech agentů, budeme místo $C_G A$ psát pouze CA.

Před definicí splňování zaveďme ještě následující značení:

- $E^0_j A$ je ekvivalentní s *A* a
- $E^{n+1}_j A$ je ekvivalentní s $EE^n_j A$.

7) Pro libovolné *s* platí: $s \Vdash C_G A$ právě tehdy, když $s \Vdash E^n_G A$ pro libovolně velké přirozené číslo *n*.

Zabývejme se nyní formálními charakteristikami sdílených a obecných znalostí. Přímo z definice splňování pro společné a obecné znalosti nyní plyne, že platí-li v nějakém světě (libovolného modelu) formule $E_G A$ (příp. $C_G A$) a $H \subseteq G$, pak v daném světě platí i $E_H A$ (příp.

[10] *E* je z anglické fráze *everybody knows*, česky: „každý ví“. Formule $E_G A$ bude značit *všichni agenti z množiny G ví, že A*.

[11] *C* je z anglického *common knowledge*, česky: „obecná znalost“. Formule $C_G A$ znamená *A je obecnou znalostí mezi agenty z množiny G*.

$C_H A$), čili společné/obecné znalosti nějaké množiny agentů jsou společnými/obecnými znalostmi i libovolné podmnožiny této množiny agentů.

Povšimněme si také, že pokud v některém světě platí $E^n_G A$, pak ve všech světech, kam se z nich po *n* krocích dostaneme po libovolné hraně značené číslicí z množiny G, platí formule *A*. Pokud v nějakém světě platí $C_G A$, pak dokonce ve všech světech, kam se z nich po libovolném počtu kroků dostaneme po libovolné hraně značené číslicí z množiny G, platí formule *A*. To nás vede k následující definici.

Definice 9

Buď G množina přirozených čísel. Řekneme, že svět *t* je *G-dosažitelný* ze světa *s*, je-li *t* i-dosažitelný ze světa *s* pro každé $i \in$ G. Řekneme, že svět *t* je *G-dostupný* ze světa *s*, existuje-li konečná posloupnost světů $s_0,...,s_n$ (*n* je libovolné) tak, že

- $s_0 = s$
- $s_n = t$
- pro každé *i* v intervalu <1, n> platí: svět s_i je G-dosažitelný ze světa s_{i-1}.

Z obou definic nyní plyne, že v libovolném světě *s* (libovolného modelu) platí formule $E_G A$ právě tehdy, když formule *A* platí v každém světě, jenž je v daném modelu ze světa *s* G-dosažitelný, a že v něm platí formule $C_G A$, právě tehdy, když formule *A* platí v každém světě, jenž je z něj G-dostupný. Z toho je patrné, že obecné znalosti nějaké množiny agentů jsou také společnými znalostmi této množiny agentů. Obecněji: Platí-li v nějakém světě libovolného modelu $C_G A$, pak v něm platí i $E^n_G A$ pro libovolné *n*.

Dále můžeme snadno nahlédnout, že jak relace G-dosažitelnosti, tak i relace G-dostupnosti jsou ekvivalencemi, jsou-li všechny relace dostupnosti příslušného modelu ekvivalencemi. Pro operátory E_G a C_G proto budou platit (mimo jiné) schémata **K**,**T**, **B**, **4**, **5**.

Podívejme se na model na obrázku 5. Čtenář si jistě povšiml, že průnik všech světů 1-dosažitelných z nějakého světa *s* a všech světů 2-dosažitelných ze stejného světa *s*, je vždy jednoprvková množina obsahující pouze svět *s*. Jinými slovy, pro každý možný svět *s* v daném modelu platí, že jediný možný svět, jenž je zároveň 1-dosažitelný i 2-dosažitelný ze světa *s*, je opět svět *s*. To je dáno tím, že mezi prvního a druhého agenta jsou distribuované veškeré znalosti o celé situaci. Kdyby měl jeden z agentů možnost podívat se i na kartu, kterou má v ruce druhý agent, už by dokázal identifikovat aktuální svět a logicky odvodit, jaký zbývající symbol je zakreslen na otočené kartě na stole. Mezi oběma agenty je tedy

distribuována veškerá znalost aktuálního rozložení všech tří karet. Na základě předchozího pozorování si nyní uveďme formální definici *distribuované znalosti*:

Definice 10 (distribuovaná znalost)

Nechť G je podmnožina množiny agentů. Zaveďme do jazyka epistemických logik nový modální operátor D_G pro distribuovanou znalost.

Formule D_GA bude znamenat *mezi agenty z množiny G je distribuovaná znalost, že A.*

Pokud G bude množina všech agentů, budeme místo D_GA psát prostě DA.

Řekneme, že nějaký svět *t* je *G-společný* světu *s*, jestliže *t* je i-dostupný ze světa s pro všechna $i \in G$.

Splňování definujeme následovně:

8) Pro libovolné *s* platí: $s \Vdash D_GA$ právě tehdy, když pro libovolný svět *t* G-společný světu *s* platí: $t \Vdash A$

Z uvedené definice je jasné, že pokud je některá znalost sdílená mezi skupinou agentů, pak má stejnou znalost každý ze skupiny agentů. Tím spíše je uvedená znalost mezi nimi distribuovaná. Pro libovolnou podmnožinu množiny agentů G a libovolnou formuli *A* pak platí (myšleno v libovolném modelu nad libovolným rámcem):

$$E_GA \rightarrow K_iA$$

a

$$E_GA \rightarrow D_GA$$

pro každé $i \in G$.[12]

Nakonec si ještě povšimněme formálních vlastností distribuované znalosti. Platí-li v nějakém světě libovolného modelu formule D_GA a $G \subseteq H$, pak v daném světě platí i D_HA. To znamená, že distribuovaná znalost se zachová i při nárůstu členů ve skupině. Platí-li *A* v celém modelu, pak platí i D_GA pro libovolnou množinu agentů G. K tomu ještě dodejme, že pro distribuovanou znalost platí (mimo jiné) všechna schémata **K**, **T**, **B**, **4**, **5**. Jak distribuovaná, tak i sdílená a obecná znalost se proto shodují na základních logických charakteristikách běžných znalostí.

[12] V modelu na obr. 8 platí i obrácené implikace $K_1A \rightarrow EA$ a $K_2A \rightarrow EA$. Jelikož z předchozího pozorování plyne i $EA \rightarrow K_1A$ a $EA \rightarrow K_2A$, pak v daném modelu platí $K_1A \leftrightarrow EA \leftrightarrow K_2A$. Dokonce platí i implikace $DA \rightarrow K_1A$ a $DA \rightarrow K_2A$, čili $DA \rightarrow EA$, a jelikož obecně platí implikace $EA \rightarrow DA$, pak zde platí i $EA \leftrightarrow DA$.

Racionální jednání a logika

Formální přístupy k racionálnímu jednání se vyznačují velkou různorodostí. To, co je skutečně spojuje, je jejich snaha zachytit racionálního aktéra, jenž zvažuje různé alternativy svého jednání. V humanitních vědách je asi nejznámějším přístupem teorie her. Epistemická logika se naproti tomu objevuje zejména v oblasti reprezentace znalostí v rámci umělé inteligence. V tomto smyslu je epistemická logika, pracující s tzv. *racionálním aktérem*, mnohem mladší disciplínou, než je tomu u teorie her. Propojení modální logiky a analýzy znalostí v podobě, jak je i dnes užívána, bychom našli na počátku 60. let 20. století. Aplikace tohoto přístupu se však objevují až v 80. letech při vývoji autonomního *software* a *hardware*.

V textu nám šlo především o jednoduché představení základů epistemické logiky. Zvolili jsme spíše populární formu motivačních příkladů, které se obvykle uvádí jako určitý vzor pro zavedení některých formálních přístupů. Text tedy poskytuje základní pojmy a definice toho, co je běžně používáno při studiu informačních toků, komunikace a sdílení znalostí.

Literatura

Běhounek, L. 2005. „Formální sémantika logiky modalit.“ In V. Kolman (ed.), *Možnost, skutečnost, nutnost.* Praha: Filosofia.

Demlová, M, Pondělíček, B. 1997. *Matematická logika.* Praha: České vysoké učení technické.

van Ditmarsch, H., van der Hoek, W., Kooi, B. 2006. *Dynamic Epistemic Logic.* Springer.

Fagin, R., Halpern, J.Y., Moses, Y., Vardi, M.Y. 1995. *Reasoning about Knowledge*. Cambridge: MIT.

Hughes, G.E., Cresswell, M.J. 1996. *A New Introduction to Modal Logic*. London: Routledge

Jirků, P., Vejnarová, J. 2005. *Formální logika.* Praha: Oeconomica, VŠE.

Kubík, A. 2004. *Inteligentní agenty: tvorba aplikačního software na bázi multiagentových systémů.* Brno: Computer Press.

Meyer, J.-J., van der Hoek, W. 1995. *Epistemic logic for AI and Computer Science.* Cambridge.

Meyer, J.-J., Veltman, F. 2007. „Intelligent Agents and Common Sense Reasoning.“ In P. Blackburn et al. (eds.), *Handbook of Modal Logic*. Elsevier.

Opava, Z. 1989. *Matematika kolem nás*. Praha: Albatros.

Peliš, M. 2007. „Teorie her jako formální teorie racionálního rozhodování.“ In J. Šubrt (ed.), *Soudobá sociologie II*. Praha: Karolinum.

Peregrin, J. 2004. *Logika a logiky*. Praha: Academia.