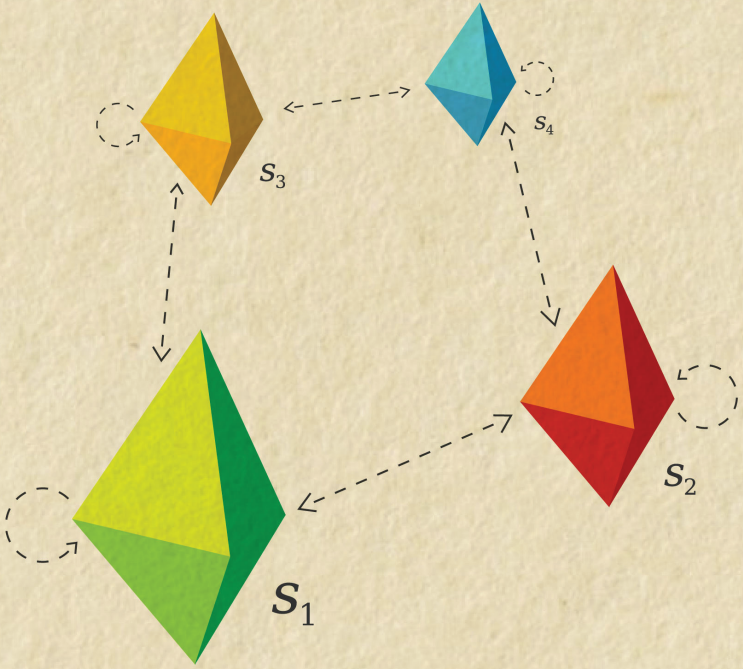


Epistemická logika:

úvod se zaměřením na studenty humanitních oborů

Ivo Pezlar

Masarykova univerzita
Brno 2015

Epistemická logika:
úvod se zaměřením na studenty humanitních oborů

Ivo Pezlar

Masarykova univerzita

Brno 2015

Publikace byla vytvořena v rámci projektu „Logika: systémový rámec rozvoje oboru v ČR a koncepce propedeutik pro mezioborová studia“, č. reg. CZ.1.07/2.2.00/28.0216 Operační program Vzdělávání pro konkurenceschopnost spolufinancovaného z Evropského sociálního fondu a státního rozpočtu České republiky.

Knihu recenzovali:
Mgr. Igor Sedlár, PhD.
RNDr. Ondrej Majer, CSc.

ISBN 756+6.+0/.+5744+1 &mljglc8nbd'
GQ@L 756+6.+0/.+557/+/ &tĆx_lĆ t_x`_'

Obsah

Seznam obrázků

Kapitola 1

Úvod

Pojem znalosti stojí v centru pozornosti západní filosofie již od jejího počátku v antickém Řecku. Co je to znalost? Jak ji můžeme zdůvodnit? Jaké jsou její hranice? Tyto a jim podobné otázky postupně vedly ke vzniku epistemologie, disciplíny zkoumající lidské vědění a poznání. Myšlenka epistemické logiky, tj. logiky zabývající se pojmem znalosti, je o poznání mladší. Lze ji vystopovat už u středověkých myslitelů jako Petr Abélard, William Ockham či Pseudo-Scotus,[1] nicméně byla to až polovina 20. století, kdy epistemická logika začala nabývat konkrétnějších obrysů. Vedle tradičních epistemologických otázek se začala objevovat i řada nových: Jak popsat znalost formálně? Jaké jsou její logické vlastnosti? apod.

V této souvislosti je nutné připomenout zejména Rudolfa Carnapa a jeho analýzu vět připisujících přesvědčení jako „Alenka je přesvědčena, že P“ (angl. belief-sentences).[2] Kniha *An Essay in Modal Logic* z roku 1951 od George H. von Wrighta má v historii epistemické logiky

[1] Srov. [Boh, 1993].

[2] Srov. [Carnap, 1947].

rovněž své nezpochybnitelné místo.[3] To stejné můžeme říci i o článku *New Foundations for Lewis Modal Systems* z roku 1957 od Johna Lemmona.[4] A rozhodně bychom měli alespoň zmínit i práce Jerzy Łośe,[5] Arthura Priora[6] či Nicholase Reschera.[7,8]

Na první důslednou, ucelenou a systematickou formální analýzu pojmu znalosti jsme si avšak museli počkat až do 60. let minulého století, kdy finský filosof a logik Jaakko Hintikka vydává svoje klíčové dílo *Knowledge and Belief* (dále jen *K&B*[9]), které prakticky položilo základy toho, co je dnes chápáno jako epistemická logika.

Hintikka sám ovšem za zakladatele moderní epistemické logiky považuje svého, výše již zmíněného, krajana von Wrighta, nicméně to nic nemění na skutečnosti, že to byla především kniha *K&B*, resp. systém v ní představený, který ustanovil epistemickou logiku jako svébytné odvětví filosofické logiky. Hector-Neri Castañeda, guatemalský filosof a zakladatel časopisu *Noûs*, jej dokonce označil za pravděpodobně nejdůležitější filosofický nástroj od objevu striktní implikace.[10]

Původní motivací Hintikkovy epistemické logiky bylo naleznutí explicitních kriterií bezespornosti pro množiny tvrzení obsahující takové výrazy jako „vím“, „věřím“, „jsem přesvědčený“ atd.[11] Tato kritéria měla

[3] Srov. [von Wright, 1951].

[4] Srov. [Lemmon, 1957].

[5] Srov. [Łoś, 1948].

[6] Srov. [Prior, 1957].

[7] Srov. [Rescher, 1960].

[8] Jistě bychom byli schopni naleznout i další autory, kteří se menší či větší měrou zabývali formálním zkoumáním znalosti a přesvědčení, nicméně naším cílem zde není poskytnout úplnou historii epistemické logiky.

[9] Srov. [Hintikka, 1962].

[10] Srov. [Castañeda, 1964].

[11] Přesněji, Hintikka nemluví o bezespornosti (konzistenci), ale o tzv. *obhajitelnosti*

následně umožnit rigorózní manipulaci s pojmy jako znalost či přesvědčení. Je nutné zmínit, že tento přístup, tj. logická analýza či obecněji sémantické zkoumání znalosti a přesvědčení, má vedle řady zastánců,[12] i své — i když dnes už spíše jen historické — odpůrce.[13]

Epistemická logika začala brzy přitahovat pozornost dalších oborů: ekonomie, teorie her, aplikované i teoretické informatiky a především pak umělé inteligence, neboť mezi logickou analýzou znalosti a její formální reprezentací je velmi tenká hranice, lze-li vůbec o nějaké mluvit. A tak původní idea „hintikkovské" epistemické logiky jakožto v podstatě pojmového zkoumání postupně ustupovala před její praktickou aplikací (např. v multiagentních systémech) kladoucí důraz spíše na souvislost znalosti s jednáním a výpočetní složitostí. Tímto směrem se ale v této knize ubírat nebudeme, na místo toho se pokusíme navázat na filosoficky motivovanou epistemickou logiku z *K&B*, tj. logiku, jejíž hlavním cílem byla adekvátní formální explikace znalosti.

V *K&B* nás Hintikka seznámil s formálním modelem znalosti, který byl vystavěný na modální intenzionální logice se sémantikou možných světů. Z epistemologického hlediska ovšem jeho systém v jistém ohledu selhal, neboť kladl příliš vysoké nároky na poznávající subjekt, resp. agenta, což vyústilo v tzv. *problém logické vševědoucnosti*. Naším hlavním cílem bude tento sémantický model znalosti představit, uvést argumenty pro jeho zavedení, ukázat jeho silné a slabé stránky a nabídnout různá řešení logické vševědoucnosti.

Nebudeme však přesně kopírovat Hintikkův systém, ale vystavíme

(angl. defensibility), srov. [Hintikka, 1962], s. 26.

[12] Srov. [van Benthem, 2006], [Hendricks, 2006], [Stalnaker, 2006], [Williamson, 2000], [Holliday, 2013].

[13] Srov. [Hocutt, 1972], [Hales, 1995].

svým způsobem „učebnicový“, standardní systém epistemické logiky. Půjde tedy o modelový příklad, který se ovšem bude držet své předlohy ve všech podstatných rysech, tj. půjde o koncept znalosti založený na modální intenzionální logice se sémantikou možných světů.

Kniha byla psána tak, aby byla přístupná každému, kdo se chce něco dozvědět o epistemické logice, ale neví kde začít. Tento text je tedy určen všem, kteří se chtějí seznámit se základy epistemické logiky a problémy, které s sebou tato disciplína přináší. S tím je spojen i sekundární cíl knihy, a to poskytnout v rámci možností srozumitelný úvod do epistemické logiky, tj. do logické analýzy znalosti. Z toho důvodu zvolíme populárnější styl výkladu, neboť — jak Hintikkovi vytýkal např. americký filosof Roderick Chisholm — *K&B* je složitější než by musela být, a její technická povaha má sklony odrazovat všechny kromě modálních logiků.[14] To je nesporná škoda, neboť *K&B* aspirovala být především epistemologickým dílem, ne jen cvičením ve formální logice.

Mnohá témata, která se v této knize objeví, patří již k folklóru epistemické logiky,[15] a proto budeme opomíjet důkazovou část a zaměříme se spíše na představení klíčových myšlenek v pozadí. Odbornější termíny užité výše (sémantika možných světů, intenzionální logika, . . .) budou postupně vysvětleny v průběhu samotné knihy, nicméně u čtenáře je předpokládána alespoň základní obeznámenost s výrokovou, predikátovou a modální logikou.[16] V práci se rovněž vyskytnou části techničtějšího rázu (především v kapitolách 5 a 6), kterým se vzhledem k povaze disku-

[14] [Chisholm, 1963], s. 773.

[15] Mezi klíčové publikace v tomto ohledu patří zejména knihy [Fagin et al., 1995] a [Meyer & Hoek, 1995].

[16] Obecné úvody do logiky v češtině lze naleznout např. v [Peregrin, 2004], [Švejdar, 2002], [Sochor, 2011], [Svoboda, 2010], [Kolman, 2005] či [Priest, 2007].

tovaného problému nelze vyhnout, avšak čtenáře, který si nepotrpí na formální záležitosti, jistě potěší, že tyto části nejsou nutné k ucelenému pochopení textu. Lze je tedy přeskočit, aniž by celkové sdělení textu výrazněji utrpělo. Nicméně vzhledem k názvu knihy autor předpokládá, že každý, kdo ji otevřel, má alespoň nestranný, když ne zrovna kladný vztah k logice. A to úplně stačí.

Struktura knihy. V kapitole 1 se stručně seznámíme s logicko-filosofickým pozadím problematiky propozičních postojů a jejich návazností na epistemickou logiku.

V kapitole 2 si představíme pět epistemických úsudků, které nám budou sloužit jako případové studie v průběhu celé knihy, a zamyslíme se nad jejich platností.

V kapitole 3 se seznámíme s extenzionálními systémy a pokusíme se analyzovat úsudky z druhé kapitoly v rámci tzv. extenzionální sémantiky.

V kapitole 4 si předvedeme intenzionálními systémy a úsudky z druhé kapitoly analyzujeme v rámci tzv. intenzionální sémantiky. Ještě předtím si však zkonstruujeme SEL neboli standardní epistemickou logiku založenou na modální intenzionální logice s Kripkeho sémantikou možných světů, kterou použijeme k jejich analýze.

V kapitole 5 se seznámíme s problémem logické vševědoucnosti a jeho několika vybranými řešeními.

V kapitole 6 shrneme dosažené výsledky a krátce se zmíníme o dalších směrech v rámci epistemické logiky.

1.1 Stručné historické pozadí

Doposud jsme používali termín epistemická logika v širším slova smyslu, který v sobě zahrnuje jak logiku znalostí (tj. epistemickou logikou v užším slova smyslu[17]), tak i logiku přesvědčení (tj. doxastickou logiku[18]). Budeme v tom pokračovat i nadále, pokud nebude řečeno jinak.[19]

Dějiny moderní logiky tradičně začínají německým matematikem a filosofem Gottlobem Fregem a ani v případě epistemické logiky neuděláme chybu, když začneme u něj. Ne že by snad Frege přímo předjímal vznik epistemické logiky, to určitě ne, nicméně otázky, na které narazil ve svém slavném článku *Über Sinn und Bedeutung* z roku 1892 jsou důležité pro pochopení problémů, se kterými se potýká epistemická logika.[20] Mluvit o epistemické logice až od roku 1962, tj. od vydání *K&B* je sice oprávněné, ale je to tak trochu jako vyprávění příběhu od půlky.

Frege se ve své stati dotkl více oblastí, nicméně nás bude zajímat především druhá část článku, v níž narazil na problematiku vedlejších vět uvozených slovesy jako „vědět“ či „domnívat se“, se kterými to — dle Fregových slov — vypadá dosti složitě. O co přibližně šlo? Mějme např. následující úsudek:

---

[17] Od řeckého slova „epistémé“ označující znalost, vědění.

[18] Od řeckého slova „doxa“ neboli domnívání.

[19] Anglický výraz „knowledge“ lze do češtiny přeložit jako „znalost“ nebo „vědění“. V této knize dáme přednost prvnímu z možných překladů, nicméně znalost od vědění nebudeme nijak rozlišovat a budeme s nimi nakládat jako se synonymy. Obdobně budeme postupovat i v případě slova „belief“, které lze přeložit jako „přesvědčení“, „domněnka“ či „víra“. Přestože se významy těchto slov v češtině poněkud liší, zde s nimi budeme opět nakládat jako se synonymy, přičemž upřednostňovat budeme termín „přesvědčení“.

[20] Srov. [Frege, 1892], [Frege, 1992].

Alenka ví, že Jitřenka je Jitřenka.
Jitřenka je Večernice.

Alenka ví, že Jitřenka je Večernice.

Frege rozpoznal, že v tomto a jemu podobných případech úsudků s tzv. *propozičními postoji*[21] nelze výraz „Jitřenka" jednoduše nahradit výrazem „Večernice", byť mezi nimi byla postulována identita, tj. Jitřenka je Večernice. To je v rozporu s předpokladem klasické extenzionální logiky, že identické výrazy mohou být vzájemně zaměněny beze změny pravdivostní hodnoty tvrzení, v němž se vyskytují (tzv. *princip substitutivity identických entit*). Nebudeme se zde ovšem zabývat Fregovým řešením, ale rovnou se zaměříme na spojitost mezi propozičními postoji a epistemickou logikou, neboť úsudek výše je typickým příkladem úsudků, kterými se zabývá právě epistemická logika.[22]

Za vznik moderní modální logiky, o níž se Hintikkův systém opírá, naopak vděčíme velkou měrou Clarenci I. Lewisovi, který představil dnes už kanonické systémy S1 až S5.[23] Další důležitou postavou, kterou je třeba zmínit, je Rudolf Carnap,[24] na kterého můžeme pohlížet jako na spojovací článek mezi Fregem a Hintikkou. Carnapovo zkoumání propozičních postojů v rámci tzv. intenzionální logiky je jedním z výchozích bodů *K&B*. Hintikka navázal na Carnapovu teorii *popisů stavů* (angl.

[21] *Propoziční postoj* je postoj agenta (individua, subjektu…) k určité propozici, resp. k tomu, co můžeme intuitivně chápat jako význam vnořeného tvrzení (věty…). Např. ve větě „Alenka věří, že sníh je bílý" je propoziční postoj zastoupen formulací „Alenka věří, že".

[22] Stojí ovšem za poznámku, že problematika propozičních postojů je svébytným filosofickým tématem a příklad s věděním výše je jen jedním z druhů propozičních postojů. Dále se můžeme např. také setkat s propozičními postoji jako chtění, přání apod. Srov. např. kapitola VII. *Propoziční a pojmové postoje* v [Svoboda, 2010] či [Raclavský, 2009].

[23] Srov. [Lewis & Langford, 1959].

[24] Srov. [Carnap, 1947].

state descriptions) a rozšířil ji zavedením *relace dosažitelnosti* (angl. accessibility relation), kterou přejal od amerického filosofa a logika Saula Kripkeho.[25] Naše epistemická logika se tedy bude opírat o kripkovskou sémantiku možných světů.

V *K&B* modeluje Hintikka znalost jako určitý epistemický ekvivalent nutnosti. Přestože lze zárodky chápání znalosti jako určité nutnosti naleznout už u Platóna a studium modalit jako nutnost či možnost (tzv. *aletické modality*) se datuje zpátky až k Aristotelovi, byla to až scholastická logika a filosofie, která se analogiemi mezi znalostí a nutností zabývala explicitně.[26] A byl to právě tento směr myšlení, na kterém Hintikka vystavěl svoji epistemickou logiku.

Hintikkovu epistemickou logiku z *K&B* můžeme zjednodušeně charakterizovat jako modální intenzionální logiku s kripkovskou sémantikou možných světů, ovšem s tím rozdílem, že modální operátor nutnosti $\Box$ je nahrazen operátorem znalosti K. Avšak je nutné zmínit, že sám Hintikka byl proti tomu, aby se epistemická logika chápala pouze jako jen další technická nástavba modální logiky, a tudíž se zde dopouštíme částečné dezinterpretace.[27] Ale to už příliš předbíháme, nyní jsou na řadě epistemické úsudky a extenzionální systémy.

[25] Srov. [Kripke, 1963].

[26] V tomto ohledu můžeme např. zmínit tvrzení tvaru „Je známo, že A, tedy nemůže platit $\neg A$“, které je příkladem tzv. epistemického čtení aletických modalit.

[27] Srov. [Hendricks, 2006], s. 140.

Kapitola 2

Epistemické úsudky

Hintikka zjednodušeně charakterizoval epistemickou logiku následujícím způsobem:

> Epistemická logika začíná studiem logického chování výrazů tvaru „A ví, že“. Jedním z hlavních cílů tohoto bádání je pak analýza dalších konstrukcí obsahujících výraz „ví/vědět“ prostřednictvím již zmíněného „A ví, že“.[1]

A tedy i my začneme s úsudky, které obsahují výrazy této formy. Prohlédněme si následující pětici úsudků:

$$\text{(1)}\ \frac{\begin{array}{c}\text{Alenka ví, že prší.}\\ \text{Jestliže prší, pak není pravda, že neprší.}\end{array}}{\text{Alenka ví, že není pravda, že neprší.}}$$

[1] Srov. [Hintikka & Halonen, 1998], s. 2. V originále: „Epistemic logic begins as a study of the logical behavior of the expression of the form 'b knows that.' One of the main aims of this study is to be able to analyze other constructions in terms of 'knows' by means of 'b knows that.'“ (překlad autor).

$$(2)\ \frac{\begin{array}{c}\text{Alenka ví, že každý člověk je smrtelný.}\\ \text{Alenka ví, že Sokrates je člověk.}\end{array}}{\text{Alenka ví, že Sokrates je smrtelný.}}$$

$$(3)\ \frac{\begin{array}{c}\text{Alenka ví, že Barack Obama je Barack Obama.}\\ \text{Barack Obama je prezidentem USA.}\end{array}}{\text{Alenka ví, že Barack Obama je prezidentem USA.}}$$

$$(4)\ \frac{\begin{array}{c}\text{Alenka ví, že Narvik leží severně od Osla.}\\ \text{Alenka ví, že } x \text{ leží severně od } y \text{ právě tehdy, když } y \text{ leží jižně od } x.\end{array}}{\text{Alenka ví, že Oslo leží jižně od Narviku.}}$$

$$(5)\ \frac{\begin{array}{c}\text{Alenka ví, že } \big((p \wedge q) \supset p\big) \text{ je teorém.}\\ \big((p \wedge q) \supset p\big) \leftrightarrow \big((q \wedge p) \supset p\big)\end{array}}{\text{Alenka ví, že } \big((q \wedge p) \supset p\big) \text{ je teorém.}}$$

Nyní se zamysleme nad úsudky (1)–(5) z hlediska jejich platnosti.

První úsudek (1) není platný, tj. závěr nevyplývá z premis. Obě dvě premisy mohou být pravdivé a závěr přesto nepravdivý. Není tak problém představit si např. takovou situaci, ve které Alenka zodpoví otázku „Prší?“ kladně, ale současně na dotaz „Je pravda, že není pravda, že neprší?“ odpoví „Ne“ či „Nevím“.

Forma úsudku (2) by nás mohla svádět k tomu, abychom mu přiřkli platnost, avšak Alenka nemusí znát závěr, přestože zná obě premisy. Jistě bychom mohli Alenku nařknout z nedůslednosti či z nedostatečného reflektování své znalosti, ale to nemění nic na té skutečnosti, že Alenka to

zkrátka nemusí vědět. Přesněji, nemusí si to uvědomit. V tomto případě bychom o ní nejspíše řekli, že si nedala „dvě a dvě dohromady“.

Úsudek (3) je rovněž problematický. Alenka jistě může znát Baracka Obamu a být přesvědčena o tom, že je identický sám se sebou, aniž by věděla, že je momentálně prezidentem USA. Jistě v případě tak známe osobnosti jakou je prezident USA je to poněkud hůře představitelné, ale stejná forma argumentu se vztahuje i na další případy. Např. Alenka může mít nějakého kamaráda Davida a nevědět to, že je předsedou klubu českých turistů apod.

Úsudek (4) bychom mohli mít opět nutkání prohlásit za platný, ovšem znovu lze uplatnit stejné námitky jako v případě úsudků (1) a (2).

Poslední úsudek (5) rovněž nelze považovat za platný. Pokud někdo ví, že $\big((p \wedge q) \supset p\big)$ je teorém, jistě to ještě neznamená, že ví i to, že $\big((q \wedge p) \supset p\big)$ je teorém, byť se jedná o ekvivalentní formule (dotyčná osoba nemusí být obeznámena s komutativitou konjunkce).

Shrnutí. Žádný z úsudků (1) až (5) nelze prohlásit za obecně platný. Vždy šlo nalézt protipříklad, který popřel platnost konkrétního odvození (inference). Vzhledem k jednoduchosti zvolených úsudků lze neúspěšná odvození vysvětlit v podstatě pouze odkazem na agentovu (v tomto případě Alenčinu) nedůslednost, nesoustředěnost, lenost anebo neobeznámenost s použitým pojmem. Důvodů, proč agent může při epistemických odvozeních selhat, je ale samozřejmě více a stejně tak se i stupňuje jejich závažnost. Při složitějších úsudcích s více premisami vstupují totiž do hry i takové faktory jako čas a paměť. Odvození závěru ze stovky premis bude jistě trvat déle a vyžadovat více paměti než odvození závěru z premis dvou atd.

Jsou tu ale vůbec nějaká epistemická odvození, která si mohou nárokovat univerzální platnost? Lze z nějakého epistemického tvrzení odvodit vždy bezpečně nějaké další epistemické tvrzení? Položíme-li otázku takto, odpověď musí znít ne.[2] Vždy totiž můžeme aplikovat taktiku použitou výše a přijít s nějakým triviálním protipříkladem: nechtělo se mu, byl líný, neuvědomil si to, trpěl amnézií apod. To ovšem nijak nepodrývá snahu epistemické logiky. Jejím cílem je podání adekvátního modelu znalosti a přesvědčení *racionálního agenta*, přičemž slovo „racionální" je zde rozhodující.

Pokud je např. agent líný a nechce se mu provádět žádné odvození, epistemická logika tu není od toho, aby jej k něčemu nutila, nemá žádné normativní ambice. Jinými slovy, epistemická logika tu není od toho, aby kárala „logicky líné" agenty. Snaží se pouze popisovat to, co je rozumné předpokládat, že by měl být racionální agent schopen odvodit, zná-li takové a takové premisy. Na druhou stranu, epistemická logika musí umět zohlednit i tyto lenošné a nepozorné agenty. Jak napsal poněkud nevybíravě, leč výstižně Robert Eberle, je to logika, která musí umět zohlednit i naprosté ignoranty, úplné pitomce a ty největší hlupáky.[3] S trochou nadsázky tak můžeme říci, že epistemická logika se řídí heslem „*hope for the best, plan for the worst*", tedy doufá v ty nejracionálnější agenty, ale připravuje se na ty nejhorší.[4]

---

[2] Stojí za zmínku, že test platnosti, který jsme zde nyní prováděli, není zdaleka žádnou novinkou. Už Pseudo-Scotus si všiml toho, že pokud platný úsudek upravíme tak, že před jednu z premis přidáme „je známo, že", stále budeme moci odvodit původní závěr, ale nebudeme moci odvodit to, že je znám. Došel tak k tomu, že je třeba zavrhnout obecný princip „jsou-li známy premisy, je znám i závěr". Jak si můžeme všimnout, ke stejnému závěru jsme dospěli i my zde, jen zhruba o 650 let později.

[3] Srov. [Eberle, 1974].

[4] Pluralitu systémů epistemických logik pak můžeme vysvětlit právě jako důsledek snahy najít kompromis mezi těmito dvěma protichůdnými požadavky. Za tento po-

Naše pětice úsudků bude sloužit právě k tomu, abychom otestovali, zda hintikkovská epistemická logika dokáže zohlednit tyto nedokonalosti agentů.

Při analýzách se ovšem neomezíme jen na hledisko platnosti, ale vezmeme v potaz také otázku přesnosti, resp. informativnosti analýz. Asi není třeba dodávat, že toto nové kriterium, na rozdíl od kriteria platnosti diskutovaného výše, vnáší do analýzy poněkud subjektivní charakter řízený našimi intuicemi (Co je ještě podstatná informace? A co už není?), nicméně obecně můžeme říci, že čím více informací analýza premis a závěrů jednotlivých úsudků zachová, tím lépe.

Analýzy nabízené jednotlivými systémy epistemických logik tak budeme posuzovat ze dvou hledisek:

A. *platnost* – Zachovala analýza neplatnost úsudku? (kriterium A)

B. *informativnost* – Zachovala analýza všechny podstatné informace? (kriterium B)

Pokud nějaký ze systémů epistemické logiky, se kterými se seznámíme v nadcházejících kapitolách, analyzuje adekvátně všech pět úsudků, tj. odhalí, že žádný z úsudků (1)–(5) není obecně platný (kriterium A), a současně zachová všechny klíčové informace (kriterium B), řekneme o něm, že je vhodný pro účely epistemické logiky. A o nějakém systému X řekneme, že jeho analýza úsudků (1)–(5) je adekvátnější než analýza systémem Y, jestliže počet adekvátně analyzovaných úsudků systémem X bude větší než počet adekvátně analyzovaných úsudků systémem Y.

střeh děkuji Jiřímu Raclavskému.

2.1 Racionální agent

Výše jsme odkazovali na tzv. *racionálního agenta*. Jak jej budeme přesněji chápat? Agentem budeme rozumět nositele znalosti. A třebaže v této knize budou hrát hlavní roli lidé, mnohdy se vyplatí agenty chápat obecněji jako cokoli, co je schopné disponovat informacemi v nejširším slova smyslu, což může zahrnovat vše od budíků, teploměrů, termostatů, mikrovlnek až po expertní systémy, počítačové programy, inteligentní protézy a nejrůznější roboty.

Jednoduchý příklad: uvažme digitální budík, který má alarm nastavený na sedm hodin ráno. Budík ovšem běží o hodinu nazpět, a začne tedy zvonit až v 8:00 hodin. Tuto situaci lze jednoduše popsat tak, že si budík jen „myslel“, že je 7:00, byť ve skutečnosti už bylo o hodinu více. Samozřejmě žádné takové sofistikované uvažování v běžném budíku neprobíhá a veškeré „přemýšlení“ je předem nastaveno (angl. tzv. hard-wired), o natahovacích budících nemluvě. Přesto ale lze chování budíku velmi snadno vysvětlit tak, že zvoní tehdy, když je přesvědčen, že aktuální čas se shoduje s nastaveným časem alarmu. Stručně řečeno, k tomu, abychom obecně nějakému agentovi připsali určitá přesvědčení, není nutné u něj předpokládat ani uvažovací schopnosti, ani vědomí. To, že agent není schopen mít žádná přesvědčení, ještě neznamená, že mu nemůžeme žádná „zvnějšku“ připsat, a vysvětlit tak např. jako chování.

Tím se dostáváme ke dvěma možným interpretacím epistemické logiky. Můžeme ji buď chápat tak, že popisuje agentovo uvažování o znalosti a přesvědčení z hlediska *první osoby*, tj. že se jedná o vnitřní jazyk, pomocí kterého agent uvažuje o světě, nebo z hlediska *třetí osoby*, kdy epistemická logika v podstatě slouží k rekonstrukci či přesněji k repre-

zentaci agentova uvažování. Např. v případě budíku a jeho předčasného zvonění byla epistemická logika vlastně užita jako metajazyk k popisu a vysvětlení dané situace. Byla to právě tato její druhá možná interpretace z externího hlediska, která přitáhla pozornost informatiků a badatelů z oblasti umělé inteligence. Přestože tato distinkce nemá žádný zásadní vliv na formální model znalosti, z hlediska obecnějšího epistemologického kontextu hraje otázka perspektivy důležitou roli. V rámci zachování Hintikkovy tradice se budeme držet první interpretace.

Představení pětice epistemických úsudků máme za sebou, stejně jako i určení jejich platnosti. Nyní se můžeme konečně vrhnout na jejich analýzu v rámci extenzionálních systémů.

Kapitola 3

Extenzionální systémy

Extenzionálními systémy budeme rozumět klasické logické systémy, ve kterých je význam určitého tvrzení (výroku) chápán jako jeho pravdivostní hodnota (tj. pravda, nebo nepravda) a které zachovávají princip kompozicionality a substitutivity identických entit. V určitém systému platí *princip kompozicionality*, pokud je pravdivostní hodnota složeného tvrzení dána pravdivostními hodnotami jeho částí a *princip substitutivity identických entit* platí tehdy, mohou-li být ekvivalentní výrazy v libovolném tvrzení zaměněny beze změny pravdivostní hodnoty celého tvrzení. Mezi extenzionální systémy patří např. klasická výroková a predikátová logika. Nejdříve se podíváme, jak si s našimi epistemickými úsudky poradí aparát výrokové logiky.

3.1 Výroková logika

Na výrokovou logiku můžeme pohlížet jako na teorii logických spojek, které jsou chápány jako pravdivostní funkce, tj. zobrazení z pravdivost-

ních hodnot do pravdivostních hodnot. V přirozeném jazyce jsou logické spojky zastoupeny výrazy jako „a“, „ne“, „nebo“, „jestliže. . . , pak. . . “, „. . . právě tehdy, když. . . “ atd. Z toho vyplývá, že pokud určité tvrzení neobsahuje žádnou z těchto spojek, je z hlediska výrokové logiky neanalyzovatelné, resp. analyzovatelné pouze triviálně.[1] Expresivita výrokové logiky je tedy velmi malá. Uvažme úsudek (1):

$$(1)\ \frac{\begin{array}{c}\text{Alenka ví, že prší.}\\ \text{Jestliže prší, pak není pravda, že neprší.}\end{array}}{\text{Alenka ví, že není pravda, že neprší.}}$$

Tento úsudek můžeme zachytit ve výrokové logice následujícím způsobem:

$$(1')\ \frac{\begin{array}{c}p\\ q \rightarrow \neg\neg q\end{array}}{\not{r}}$$

Na první pohled by se mohlo zdát, že máme vyhráno. Z p a $q \rightarrow \neg\neg q$ rozhodně nevyplývá r, úsudek je tudíž neplatný (značíme přeškrtnutým závěrem r, tj. $\not{r}$), což je v souladu s tím, k čemu jsme dospěli v předchozí kapitole.

Z hlediska platnosti se tedy jedná o korektní analýzu. Ovšem je to analýza velmi povrchní, při které se navíc ztratilo hned několik informací. Čtenář jistě vidí, že premisy a závěr úsudku (1) spolu určitým způsobem souvisejí, avšak tuto informaci nám naše analýza ve výrokové logice nezachovala. Dalo by se říci, že k žádné analýze vlastně ani nedošlo, jen jsme nahradili jednotlivé výroky symboly p, q, r a logické spojky

[1] K přesnějšímu vymezení výrokové logiky se dostaneme v sekci 4.2.

„jestliže…, pak…“ a negaci („není pravda, že…“, „ne–…“) symboly $\rightarrow$ a $\neg$.[2]

Přestože tedy tato „analýza“ adekvátně zachycuje neplatnost úsudku (1) (viz kriterium A v předchozí kapitole), činí tak na úkor jeho informativnosti (kriterium B), a proto ji nemůžeme považovat za uspokojivou.

Přistupme nyní k úsudku (2):

$$(2)\ \frac{\begin{array}{c}\text{Alenka ví, že každý člověk je smrtelný.}\\ \text{Alenka ví, že Sokrates je člověk.}\end{array}}{\text{Alenka ví, že Sokrates je smrtelný.}}$$

Premisy ani závěr neobsahují žádnou z výše zmíněných spojek, tedy jedinou možností, jak tento úsudek analyzovat v rámci výrokové logiky, je:

$$(2')\ \frac{\begin{array}{c}p\\ q\end{array}}{r}$$

Jak můžeme vidět, zde je problém diskutovaný výše ještě zřetelnější. Úsudek je sice opět správně zachycený jako neplatný, ale ztratilo se příliš mnoho údajů. Bude tedy potřeba poohlédnout se po expresivnějším systému než je výroková logika, který nám umožní přesnější analýzu úsudků.

3.2 Predikátová logika

Predikátová logika je logickou teorií kvantifikátorů, pro které v přirozeném jazyce užíváme výrazy jako „každý“, „všichni“, „někdo“, „nikdo“

[2] Informativnost analýzy bychom mohli alternativně posuzovat i z hlediska toho, co můžeme nazvat intuitivním vyplýváním. Např. z „Alenka ví, že prší“ intuitivně vyplývá „Prší“, ale ani tuto skutečnost není výroková logika schopna zachytit.

atd., vlastností individuí (objektů, předmětů, ...) a vztahů mezi nimi, přičemž tyto vztahy (relace) jsou predikovány jednotlivým individuím.[3]

Predikátová logika, na rozdíl od logiky výrokové, umožňuje analyzovat vnitřní strukturu tvrzení. Predikátovou logiku proto můžeme chápat jako zobecnění či rozšíření logiky výrokové. Vraťme se nyní zpátky k úsudku (2):

$$(2)\ \frac{\begin{array}{c}\text{Alenka ví, že každý člověk je smrtelný.}\\ \text{Alenka ví, že Sokrates je člověk.}\end{array}}{\text{Alenka ví, že Sokrates je smrtelný.}}$$

a zkusme jej analyzovat v rámci predikátové logiky. Alenka a Sokrates jsou nepochybně nějaká individua (zapíšeme jako ALENKA a SOKRATES), vlastnosti člověk a smrtelný (resp. být člověk a být smrtelný) analyzujeme jako unární predikáty (tj. x je člověk, x je smrtelný; zapíšeme jako ČLOVĚK(x) a SMRTELNÝ(x)), vědění zkusíme zachytit jako binární predikát (tj. x ví, že platí nějaké y; zapíšeme jako VĚDĚT(x, y)) a výskyt výrazu ‚každý' zaneseme do naší analýzy pomocí obecného kvantifikátoru $\forall$. Dáme-li to celé dohromady, získáme:

$$(2'')\ \frac{\begin{array}{c}\text{VĚDĚT}\Big(\text{ALENKA}, \forall x\big(\text{ČLOVĚK}(x) \rightarrow \text{SMRTELNÝ}(x)\big)\Big)\\ \text{VĚDĚT}\big(\text{ALENKA}, \text{ČLOVĚK}(\text{SOKRATES})\big)\end{array}}{\text{VĚDĚT}\big(\text{ALENKA}, \text{SMRTELNÝ}(\text{SOKRATES})\big)}$$

Výše jsme si ale řekli, že predikátová logika pracuje s individui a relacemi mezi nimi. To znamená, že můžeme predikovat pouze o individuích a o ničem jiném. Tento přístup je ale analýzou výše jasně porušen, jelikož

[3] K přesnějšímu vymezení predikátové logiky se dostaneme v sekci 4.4.

už v první premise je binární predikát VĚDĚT aplikován na *individuum* ALENKA a *formuli* $\forall x\big(\text{ČLOVĚK}(x) \rightarrow \text{SMRTELNÝ}(x)\big)$. Tato analýza tedy není v rámci standardní predikátové logiky přípustná. K analýze tedy budeme muset zkusit přistoupit jiným způsobem.

Pravděpodobně nejjednodušší způsob, jak se této komplikaci vyhnout, je ignorovat fráze jako „vědět, že“, resp. přesunout je z objektového jazyka do meta-jazyka. Jinými slovy, tvrzení jako:

(2a) Alenka ví, že každý člověk je smrtelný.

se nebudeme pokoušet analyzovat jako:

(2a') $\text{VĚDĚT}\Big(\text{ALENKA}, \forall x\big(\text{ČLOVĚK}(x) \rightarrow \text{SMRTELNÝ}(x)\big)\Big)$

ale jednoduše jen jako:

(2a") $\forall x\big(\text{ČLOVĚK}(x) \rightarrow \text{SMRTELNÝ}(x)\big)_{\textit{Alenka\_Ví}}$

přičemž *Alenka_Ví* je jen dodatečná poznámka, která není součástí konkrétního logického systému. Jinými slovy, ono „vědět, že“ (resp. znalost) budeme reprezentovat jen na meta-úrovni množinou predikátových formulí s dodatkem *Alenka_Ví*.

Tento přístup nám pak umožňuje analyzovat úsudek (2) jako:

$$(2") \quad \frac{\begin{array}{c}\forall x\big(\text{ČLOVĚK}(x) \rightarrow \text{SMRTELNÝ}(x)\big)_{\textit{Alenka\_Ví}} \\ \text{ČLOVĚK}(\text{SOKRATES})_{\textit{Alenka\_Ví}}\end{array}}{\cancel{\text{SMRTELNÝ}(\text{SOKRATES})_{\textit{Alenka\_Ví}}}}$$

Dobrá zpráva je, že tato analýza zachovává všechny důležité informace (s přihlédnutím k meta-úrovni). A vzhledem k tomu, že nemáme žádné

meta-pravidlo pro distribuci poznámky *Alenka_Ví* mezi premisami a závěrem v platném úsudku, ani závěr SMRTELNÝ(SOKRATES)$_{Alenka\_Ví}$ nelze považovat za obecně korektní.

Špatnou zprávou je ovšem to, že analýza úsudku (2") se spoléhá na problematické syntaktické pojetí znalosti (tj. znalost jako formule označená pomocí poznámky *Alenka_Ví*), které bude diskutované v následující sekci.[4]

Povedlo se nám tedy splnit kriteria A a B, ale za cenu toho, že jsme z konceptu znalosti udělali pouhý index na formulích, což rozhodně nelze považovat za uspokojivé řešení, obzvláště v rámci epistemické logiky. Zdá se tedy, že ani predikátová logika nebude stačit a že frázi „vědět, že" nepůjde jednoduše ignorovat (resp. přesunout na meta-úroveň).

3.3 Shrnutí výsledků

Extenzionální logika (konkrétně výroková a predikátová) selhává v zachycení jednotlivých úsudků v důsledku neschopnosti vypořádat se adekvátně s propozičními postoji. Konkrétněji, ke komplikacím dochází, když se pokoušíme zachytit frázi „vědět, že" (resp. odpovídající propoziční postoj „x ví, že platí y") jako binární predikát. To je jistě důležité zjištění. Ovšem mnohem důležitější je vědět, *proč* přesně k tomuto selhání dochází.

Příčinou jsou právě extenzionální systémy samotné, přesněji dva z jejich základních principů, a to princip *kompozicionality* a *substitutivity*

[4] Nemluvě o tom, že tento přístup neumožňuje ani zachytit řadu zajímavých vlastností, které bychom znalosti obecně připsali. Např. nejsme schopni zachytit meta-znalost, tj. znalost o nějaké znalosti (např. „Alenka ví, že něco ví") apod.

identických entit.

Nejdříve uvažme princip kompozicionality. Mějme následující tvrzení s propozičním postojem:

(e) Alenka věří, že existuje prezident USA.

Princip kompozicionality nám říká, že pravdivostní hodnota celého tvrzení je dána pravdivostními hodnotami jeho částí. Tento princip je zde ale porušen, neboť celé tvrzení může být pravdivé či nepravdivé bez ohledu na pravdivostní hodnotu vnořeného tvrzení „prezident USA existuje“. Jinými slovy, i kdyby prezident USA neexistoval, Alenka tomu stále může věřit.

Nyní uvažme princip substitutivity identických entit. Současným prezidentem USA je Barack Obama, takže můžeme přidat další tvrzení:

(k) prezident USA = Barack Obama

Princip substitutivity identických entit nás pak opravňuje k tomu, abychom mohli odvodit:

(k’) Alenka věří, že existuje Barack Obama.

Alenka ovšem nemusí věřit tomu, že Barack Obama existuje, jednoduše z toho důvodu, že nemusí vědět, že prezident USA a Barack Obama je jedna a tatáž osoba (viz (k)). Prezidenta USA a Baracka Obamu tak nelze libovolně zaměnit, a tím je porušen druhý ze základních předpokladů extenzionálních systémů, tj. substitutivita identitických entit. To nás staví před následující rozhodnutí: buď zcela opustit doménu extenzionálních systémů, nebo se pokusit tuto nesnáz nějak vyřešit v jejich rámci. Vzhledem k tomu, že první možnost se zdá poněkud drastická, začneme tou druhou.

Jednou z možností, jak obhájit extenzionální systémy, je revize samotného pojetí propozičních postojů. Chceme vlastně dosáhnout toho, aby význam vnořených tvrzení, tj. onoho „existuje prezident USA“ v „Alenka věří, že existuje prezident USA“, nebyl pojímán jako pravdivostní hodnota. Jedině tak můžeme zachovat platnost principu kompozicionality. Jak toho můžeme docílit?

Abychom na to dokázali odpovědět, nejprve si musíme položit následující otázku: co je předmětem propozičních postojů? Jinými slovy, k čemu vlastně zaujímáme postoj v propozičních postojích? Už víme, že to nemohou být pravdivostní hodnoty pojímané jako význam, jinak bychom opět skončili ve střetu s principem kompozicionality. Ale co by to tedy mohlo být?

První odpověď, která se nabízí, je, že předmětem propozičních postojů jsou zkrátka vedlejší věty, resp. jejich jména. Předpokládejme tedy, že to, čemu Alenka věří, jsou vedlejší věty uvozené spojkou „že“. Jinak řečeno, předmětem propozičních postojů je řetězec znaků:

(s_{cz}) ‚existuje prezident USA‘.

Tím zachováme nejen princip kompozicionality, ale rovněž i zamezíme neoprávněným epistemickým záměnám v důsledku principu substitutivity identických entit, neboť věta ‚existuje prezident USA‘ je nepochybně odlišná od věty ‚existuje Barack Obama‘, tj. nejsou to identické řetězce symbolů.

Důsledky této koncepce jsou ovšem problematické. Toto řešení totiž vede k závěru, že předmětem Alenčina přesvědčení je věta (s_{cz}). To můžeme vyjádřit jako:

(e’) Alenka věří větě ‚existuje prezident USA‘.

Základem tohoto přístupu je tedy předpoklad, že to, k čemu se vztahuje Alenčina znalost, není význam vedlejší věty, ale samotná vedlejší věta ‚existuje prezident USA' jakožto konkrétní řetězec znaků. Slabinou tohoto pojetí propozičních postojů je tak přílišná závislost na konkrétním jazyku. Jinými slovy, tento přístup nás vlastně zavazuje k jazykovému, ba přímo k syntaktickému pojetí znalosti. Co tím máme na mysli? Pokud Alenka umí anglicky, mohla by svoje přesvědčení vyjádřit větou:

(s_{en}) ‚the president of USA exists'.

A to je rozhodně jiná věta (jiný řetězec znaků) než věta (s_{cz}). To by ale znamenalo, že Alenka může věřit větě ‚existuje prezident USA', aniž by současně věřila větě ‚the president of USA exists', a to i za předpokladu, že ovládá oba jazyky. Bilingválně vzdělaná Alenka tak může v rámci této koncepce bezrozporně věřit větám ‚existuje prezident USA' a ‚the president of USA does not exist', a to je velmi problematický důsledek.[5] Jak se zdá, onomu drastickému kroku vedoucímu k zavržení extenzionálních systémů se skutečně nebudeme moci vyhnout.

3.4 Od extenze k intenzi

Obecně se má za to, že nutnou podmínkou adekvátní analýzy propozičních postojů, a tedy i epistemických úsudků, je přesun do tzv. intenzionálních systémů. Zavržením extenzionálních a přijetím intenzionálních systémů máme na mysli přesun z tzv. *extenzionálních* (transparentních,

[5] Návrh jak se vypořádat s propozičními postoji, který je založený na tomto pojetí znalosti, předložil např. [Quine, 1964], viz Dodatek C.

přímých) *kontextů* do tzv. *kontextů intenzionálních* (neprůhledných, nepřímých). Jak záhy zjistíme, rozdíl mezi nimi už dobře známe. Uvažme následující dvě tvrzení:

(p) Alenka má ráda kočky.

(p') Alenka má ráda čtyřnohé savce, kteří mňoukají.

Tato dvě tvrzení jsou z hlediska extenzionálních systémů ekvivalentní. To znamená, že v tvrzení (p) můžeme nahradit výraz „kočky" výrazem „čtyřnozí savci, kteří mňoukají" a pravdivostní hodnota celého tvrzení se nezmění. Tvrzení (p) a (p') jsou tedy příkladem extenzionálních kontextů, tj. kontextů, které umožňují (díky principu substitutivity identických entit) záměnu ekvivalentních výrazů *salva veritate*, tedy při zachování pravdivosti.

Uvažme však další dvě tvrzení:

(q) Alenka ví, že má ráda kočky.

(q') Alenka ví, že má ráda čtyřnohé savce, kteří mňoukají.

Tvrzení (q) a (q'), jak už víme z dřívějška, jsou příkladem propozičních postojů.[6] Výraz „kočky" zde nemůžeme zaměnit výrazem „čtyřnozí savci, kteří mňoukají" jednoduše proto, že Alenka např. nemusí vědět, že kočky jsou savci. Jinými slovy, z tvrzení:

(q) Alenka ví, že má ráda kočky.

nemůžeme odvodit:

[6] Jen pro připomenutí, propoziční postoj se obvykle chápe jako vztah agenta k významu určitého tvrzení (věty, ...), přičemž za význam tvrzení se obecně považuje propozice tímto tvrzením označovaná.

(q’) Alenka ví, že má ráda čtyřnohé savce, kteří mňoukají.

Alenka tedy může vědět, že má ráda kočky, aniž by věděla, že má ráda čtyřnohé mňoukající savce. To odpovídá i naším intuicím ohledně znalosti. Princip substituce ekvivalentních výrazů, jenž platí v extenzionálních kontextech, tedy selhává v intenzionálních kontextech.

Můžete shrnout, že extenzionální logika selhává jako epistemická logika, protože není schopna adekvátně analyzovat úsudky s propozičními postoji. Toto selhání je způsobeno tím, že extenzionální logiky nakládají s významem tvrzení (včetně tvrzení vnořených) jako s pravdivostními hodnotami. V přímých kontextech to nezpůsobuje žádné potíže, ale problém nastává v nepřímých kontextech, které jsou právě klíčové při analýze epistemických úsudků. Ukázali jsme si, že propoziční postoje se nemohou vztahovat ani k pravdivostním hodnotám, ani k větám samotným. Dalším adeptem na pozici předmětu propozičních postojů budou propozice pojímané jako intenze. Více o nich si ale povíme až v následující kapitole věnované intenzionálním systémům.

Kapitola 4

Intenzionální systémy

V předchozí kapitole jsme si ukázali, proč se extenzionální systémy nejsou schopny vypořádat s úsudky, které obsahují propoziční postoje. Po intenzionálních systémech budeme chtít, aby uspěly přesně tam, kde extenzionální systémy selhaly, tj. při určování platnosti úsudků (1) až (5).

Avšak ještě před tím, než se k nim budeme moci navrátit, si musíme představit systém, který použijeme k jejich analýze. V pořadí tak již třetím systémem explikace (po výrokové a predikátové logice) se stane modální intenzionální logika s Kripkeho sémantikou možných světů, kterou nazveme standardní epistemickou logikou (dále jen SEL). Ze všeho nejdříve se ale blíže seznámíme s intenzionálními systémy (sekce 4.1.1), myšlenkou modelování znalosti jako určité nutnosti (sekce 4.1.2), na které celá SEL stojí, a teorií možných světů (sekce 4.1.3).

4.1 Pozadí SEL

4.1.1 Intenzionální systémy: bližší pohled

Intenzionálními systémy budeme rozumět takové logické systémy, v nichž je význam určitého tvrzení pojímán jako propozice označovaná tímto tvrzením (větou...). Jak budeme chápat propozici neboli intenzi? Nejdříve si připomeňme, co je to extenze. Mějme nějaké tvrzení, např.

(h) Sněžka je nejvyšší hora České republiky.

Extenzí tohoto tvrzení (h) je pak jeho pravdivostní hodnota, tj. v tomto konkrétním případě pravda.

Jistě je ale představitelný i takový stav věcí, ve kterém Sněžka není nejvyšší horou. Jinými slovy, dokážeme si představit možné světy, ve kterých tvrzení (h) není pravdivé. Tato úvaha, že vedle toho, jak se věci ve světě skutečně mají (tzv. aktuální svět), tu existuje ještě množství dalších způsobů, jak by se věci mohly mít (tzv. možné světy), je hlavní myšlenkou v pozadí intenzionálních systémů.[1]

Možné světy budeme zjednodušeně reprezentovat jako množiny atomických formulí, které jsou v nich pravdivé. Místo termínu „možné světy" bychom tedy klidně mohli používat i neutrálnější výrazy jako např. stavy věcí, situace, scénáře apod.[2] Z tohoto pojetí dále vyplývá, že naše reprezentace možných světů bude relativní vzhledem k použitému jazyku

[1] Idea možných světů je obecně připisována německému filosofovi a matematikovi Gottfriedu Leibnizovi z přelomu 17. a 18. století, nicméně k jejímu širšímu uplatnění dochází až s nástupem modálních logik ve 20. století.

[2] Nepůjde nám tedy o filosofickou explikaci pojmu možného světa, která je zcela svébytným a hojně diskutovaným tématem (viz např. [Kripke, 1980], [Lewis, 1986], [Plantinga, 1974], [Stalnaker, 1984]), ale spíše jen o technickou reprezentaci možných světů.

(resp. jeho fragmentu). Množina všech možných světů pak bude potenční množinou množiny atomických formulí zvoleného jazyka (jeho fragmentu).

Co je tedy propozicí neboli intenzí tvrzení (h)? Je to množina všech možných světů, v nichž je tvrzení (h) pravdivé (resp. množina množin obsahující (h)).

Vraťme se např. k našemu tvrzení (h). Předpokládejme, že tu máme nějaké tři možné světy s_1, s_2, s_3, přičemž první dva světy s_1, s_2 obsahují tvrzení (h), zatímco svět s_3 tvrzení (h) neobsahuje. Jinými slovy, v prvních dvou světech je Sněžka nejvyšší horou České republiky, ve třetím nikoli. Z toho, co jsme si řekli výše, dále vyplývá, že extenzí tvrzení (h) v s_1 je pravda, v s_2 rovněž pravda a ve s_3 nepravda. Intenze, propozice neboli význam tvrzení (h) je pak tvořen světy s_1 a s_2, tj. množinou $\{s_1, s_2\}$.

Jednotlivé možné světy tedy budeme chápat jako množiny atomických formulí, které v nich platí. Z toho dále vyplývá, že nemohou existovat dva naprosto identické možné světy, neboť vždy se musí lišit alespoň jednou atomickou formulí. Pokud se dva možné světy skládají ze stejných atomických formulí, pak se jedná o jeden a tentýž svět.

Čím se od sebe ovšem možné světy lišit nemohou, jsou logické pravdy. Předpoklad je totiž ten, že ty platí ve všech možných světech. Jinými slovy, není možné, aby v některém z možných světů platilo $((p \wedge q) \supset p)$ a v jiném nikoli. Můžeme tedy zpřesnit dřívější formulaci a říci, že jednotlivé možné světy se od sebe liší *empirickými* tvrzeními (jako např. (h)), která v nich platí.

Můžeme shrnout, že intenzionální systémy na rozdíl od systémů extenzionálních nepojímají význam tvrzení jako jeho pravdivostní hodnotu (extenzi), ale jako množinu těch možných světů, ve kterých toto tvrzení

platí (tj. intenzi, resp. propozici).

4.1.2 Znalost jako nutnost

Proč vůbec modelovat znalost jako nutnost, tj. za pomoci právě modální logiky? Jak už jsme zmínili, SEL stojí na předpokladu, že mezi nutností a znalostí je jistá analogie, nyní ji prozkoumáme blíženi. Uvažme jednoduchý úsudek:

$$(\mathrm{t}_1)\ \frac{\text{Je nutné, že } p}{p}$$

Slovy: pokud je nutné, že p, pak p. To je obecně neproblematické odvození. Nyní uvažme epistemickou variantu této dedukce:

$$(\mathrm{t}_2)\ \frac{\text{Alenka ví, že } p}{p}$$

Jinými slovy, pokud Alenka zná nějaké tvrzení p, pak z toho můžeme usoudit i to, že p je pravdivé (platné), neboť pravdivost je obecně pokládána za nutnou podmínku znalosti (tj. nelze znát něco, co není pravda). Zde máme konkrétní příklad jisté analogie.[3]

Uvažme další příklad úsudku:

$$(\mathrm{t}_3)\ \frac{\begin{array}{c}\text{Je nutné, že } p\\ \text{Je nutné, že jestliže } p\text{, pak } q\end{array}}{\text{Je nutné, že } q}$$

či formálněji:

[3] Jak později zjistíme, toto odvození se opírá o implicitní předpoklad pravdivosti axiomu **(T)**.

$$\text{(mMP)}\ \frac{\begin{array}{c}\Box p \\ \Box(p \to q)\end{array}}{\Box q}$$

Slovy: pokud je nutné p a je nutné $p \to q$, pak můžeme odvodit i to, že je nutné q.[4] V rámci standardní modální logiky je tento úsudek platný. Ovšem při epistemickém čtení tomu už tak být nemusí. Tedy úsudek jako např.:

$$(\mathrm{t}_4)\ \frac{\begin{array}{c}\text{Alenka ví, že prší.} \\ \text{Alenka ví, že jestliže prší, pak není pravda, že neprší.}\end{array}}{\text{Alenka ví, že není pravda, že neprší.}}$$

či formálněji:

$$\text{(eMP)}\ \frac{\begin{array}{c}Kp \\ K(p \to \neg\neg p)\end{array}}{K\neg\neg p}$$

jehož variantu (2) známe již z kapitoly 3, je intuitivně neplatný, jelikož Alenka si nemusí uvědomit všechny logické důsledky svých znalostí.[5] Zkrátka řečeno, z toho, že známe obě premisy platného odvozovacího pravidla, nevyplývá, že známe i závěr.

Zde se nám tedy začala naše analogie mezi nutností a znalostí poněkud bortit. Ovšem ne všichni badatelé se nechali tímto selháním odradit. Von Wright trval na tom, že analogie mezi nutností a znalostí je udržitelná a že každý aletický princip C. I. Lewisova systému S4 může být transformován na platný epistemický princip. A cílem Hintikkovy knihy

[4] Jak později také zjistíme, toto odvození je instancí axiomu **(K)**, resp. tohoto axiomu formulovaného jako modální varianta odvozovacího pravidla modus ponens **(MP)**.

[5] Nutno ovšem dodat, (eMP), tj. epistemická varianta modálního **(MP)**, se v rámci epistemických logik obecně považuje za platný princip, více o tom však později.

K&B nebylo vlastně nic jiného než obhájit právě tuto epistemickou interpretaci Lewisova modálního systému S4, resp. převést modální aletické axiomy do epistemického kontextu.

Proč tolik pozornosti přitáhla právě modální logika S4? Důvodem bylo to, že její axiomy velice přesně postihovaly nejen vlastnosti, které asociujeme s pojmem nutnosti (k tomu byly koneckonců navrženy), ale také některé z našich základních intuic o znalosti. Uvážíme-li ale dosavadní vývoj epistemické logiky, není to zdaleka nic překvapujícího. Pokud z něj lze totiž vůbec něco vyčíst, je to právě to, že způsob, jakým chápeme znalost, má velmi blízko k tomu, jak pojímáme nutnost.

Nyní se blíže podíváme na možné světy, které jsme využili při popisu intenzionálních systémů výše.

4.1.3 Možné světy v epistemické logice

V předchozí sekci jsme si vymezili možné světy jako množiny atomických formulí, které v nich platí. Nyní si na několika příkladech představíme, jak můžeme teorii možných světů využít při analyzování epistemických situací.

4.1.3.1 Příklad 1: Počasí

Alenka se prochází po ulicích Brna, kde zrovna prší, přičemž nemá žádné informace o tom, jaké je zrovna počasí v Praze. Alenka tedy ví, že v Brně prší, ale neví, jestli je tomu tak i v Praze.

To, že Alenka ví, že v Brně prší, můžeme vyjádřit v termínech teorie možných světů následujícím způsobem: o Alence řekneme, že ví, že v Brně prší, právě tehdy, když ve všech světech, které Alenka v daném

okamžiku považuje za možné (tj. slučitelné s její dosavadní znalostí), není v Brně slunečno. Alenka nepovažuje za možné, že by bylo v Brně slunečno, protože má smyslovou evidenci o opaku, tj. že prší.

Jak je to ale s počasím v Praze? Vzhledem k tomu, že Alenka nemá žádné zprávy o stavu tamního počasí, považuje v základním případě[6] za možné dva světy: ten, ve kterém v Praze také prší, a ten, ve kterém je v Praze zrovna slunný den.[7] Analogicky, pokud by se Alenka dozvěděla, že v hlavním městě ČR také prší, přestala by považovat za možné ty světy, ve kterých je v Praze slunečno.

Předpokládejme však, že se k ní žádná taková informace nedostala, a co víc, Alenka začala uvažovat i o tom, jaké je asi počasí v Ostravě. K popisu takovéto situace, resp. možných kombinací hodnot prší/neprší nad Prahou a Ostravou, si už nevystačíme pouze se dvěma možnými světy, ale budeme potřebovat rovnou čtyři: svět, ve kterém prší jak v Praze, tak i v Ostravě; svět, ve kterém prší v Praze, ale ne v Ostravě; svět, ve kterém prší v Ostravě, ale ne v Praze a nakonec svět, ve kterém neprší ani v jednom z těchto měst.[8]

4.1.3.2 Příklad 2: Testová otázka

Uvažme další příklad. Alenka píše test, ve kterém má na výběr ze tří možných odpovědí A, B a C, přičemž správná je vždy jen jedna. Předpokládejme, že došla k otázce s následujícím zněním:

[6] Tedy za předpokladu, že možné světy budujeme nad množinou {prší v Brně, prší v Praze}.

[7] Alenčina množina možných světů tedy vypadá následovně {{prší v Brně}, {prší v Brně, prší v Praze}}.

[8] Obecně platí, že pokud bude agent přemýšlet nad pravdivostní hodnotou n tvrzení, bude muset uvažovat o 2^n možných epistemických alternativách. Čím méně toho bude Alenka vědět, tím více světů bude považovat za možné.

Kdo je autorem vědeckofantastického dramatu *R.U.R.*?

A. Isaac Asimov

B. Stanisław Lem

C. Karel Čapek

Nejprve předpokládejme, že Alenka zná správnou odpověď. Co to přesně znamená? Kdy Alenka zakroužkuje správnou odpověď? Zjevně právě tehdy, když bude vědět, že to nemůže být jinak, jinými slovy, když nebude mít žádné pochybnosti. V opačném případě, tj. pokud by si nebyla jistá, bychom o ní nemohli říci, že to ví. Samozřejmě by Alenka mohla jen tipovat a do správné odpovědi se strefit náhodou, ale pak bychom o ní opět těžko mohli říci, že znala správnou odpověď. Alenka tedy zakroužkuje správnou odpověď právě tehdy, když bude vědět, že to nemůže být jinak.

Znalost můžeme tedy chápat jako minimalizaci (či v ideálním případě naprostou eliminaci) pochybností a nejistot, ovšem ne v psychologickém slovo smyslu, ale spíše v technickém jako zmenšování množiny možných alternativ „jak by to mohlo být“. Alenka tedy ví, že platí nějaké x právě tehdy, když nepovažuje za možné, aby to x nebylo. Obecněji, o agentovi řekneme, že ví, že platí nějaké x pravě tehdy, když je x pravdivé ve všech světech, které považuje daný agent za možné. To se zdá jako rozumný předpoklad, neboť pokud by Alenka považovala za možné, že Stanisław Lem napsal *R.U.R.* a současně připouštěla i tu možnost, že to byl Karel Čapek (tj. nebyla by si jistá), rozhodně bychom o ní neřekli, že ví, kdo napsal *R.U.R.*

Nyní uvažme situaci, kdy si Alenka není správnou odpovědí úplně jistá. Ví, že to určitě nebyl Isaac Asimov, takže tuto možnost okamžitě vy-

loučí. Stále jí ovšem zbývají dvě možnosti (dva světy), které považuje za stejně možné: svět, ve kterém je autorem *R.U.R.* Stanisław Lem, a svět, ve kterém *R.U.R.* napsal Karel Čapek. Neví, kterou možnost zakroužkovat, protože z jejího pohledu jsou oba dva světy stejně pravděpodobné, tj. jeden není v Alenčiných očích „možnější" než ten druhý. Jinými slovy, jsou pro Alenku *epistemicky nerozlišitelné*, a tudíž neví, pro který z nich se rozhodnout. Takovýmto světům, které jsou vzájemně epistemicky nerozlišitelné, budeme říkat *epistemické alternativy*.

Pět minut před odevzdáváním testu si Alenka nakonec vzpomene, že *R.U.R.* napsal český autor. Už předtím věděla, že Stanisław Lem je z Polska a že Karel Čapek se narodil na území České republiky, takže si nakonec odvodí, že *R.U.R.* musel napsat Karel Čapek. Alenka zakroužkuje možnost C, protože v žádném ze světů, které momentálně považuje za možné, není představitelné, že by *R.U.R.* napsal někdo jiný než Karel Čapek.

4.1.3.3 Příklad 3: Sudoku

V dalším příkladě Alenka luští sudoku[9] (viz obr. 4.1):
Předpokládejme, že zrovna řeší políčko i-VII označené „?". Už při prvním pohledu může Alenka vyřadit čísla, která obsahuje daný čtverec, tj. 2, 5, 7, 8, 9. Políčko i-VII tedy může obsahovat pouze čísla 1, 3, 4 nebo 6. To máme dohromady čtyři možné alternativy. Pokud Alenka projde sloupec i, který obsahuje čísla 1, 3 a 6, zjistí, že i-VII už může obsahovat pouze

[9] Sudoku je logická hra, jejímž cílem je doplnit chybějící čísla 1–9 do předvyplněné tabulky rozdělené na 9×9 políček, která jsou seskupena do 9 čtverců (3×3). Tabulku je třeba doplnit tak, aby v každé řadě, v každém sloupci a v každém z devíti čtverců byla vždy použita všechna čísla 1 až 9, přičemž čísla se nesmí opakovat v žádném sloupci, řadě nebo v malém čtverci.

a	b	c	d	e	f	g	h	i	
5	3			7					I
6			1	9	5				II
	9	8					6		III
8				6				3	IV
4			8		2			1	V
7				3				6	VI
	6					2	8	?	VII
			4	1	9			5	VIII
				8			7	9	IX

Obrázek 4.1: Sudoku

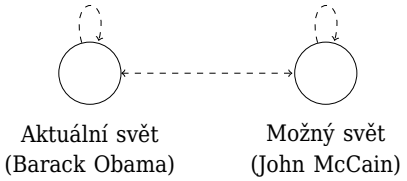
jediné číslo, a to 4. Alenka logickým uvažováním redukovala čtyři možné světy pro políčko i-VII na jeden možný svět, a tedy dospěla ke znalosti. Alenka ví, že tam bude číslo 4, protože ví, že tomu nemůže být jinak. Jinými slovy, nedokáže si představit takový svět s informacemi, jež má k dispozici (tj. tímto konkrétním zadáním sudoku a jeho pravidly), ve kterém by tomu mohlo být jinak.

4.1.3.4 Příklad 4: Vyšetřování

V tomto případě si Alenka hraje na detektiva a snaží se vypátrat, kdo jí snědl její oblíbené čokoládové sušenky. Má dva podezřelé, bratry Bedřicha a Cyrila, a na místě činu nalezla dvě stopy: obal od žvýkaček, které má rád v celé rodině jen Bedřich, a šmouhu od barvy, která naopak naznačuje přítomnost malíře Cyrila. Ani jednomu z nich ovšem nemůže krádež dokázat, neboť důkazy, které zatím nashromáždila, nejsou usvědčující. Alenka tedy uvažuje o dvou, resp. třech možných vinících: buď to provedl Bedřich, nebo Cyril, anebo oba dohromady. Tyto tři alternativy odpovídají opět třem možným světům a Alenčiným cílem je odhalit, který

z nich popisuje aktuální svět, resp. skutečný stav věcí. Problémem je, že její důkazy stejnou měrou podporují všechny tři možné světy, přesněji epistemické alternativy. Jinými slovy, jsou pro ni opět epistemicky nerozlišitelné, a dokud neobjeví další stopy, tak takovými i zůstanou.

Uveďme si další podobný příklad. Prezidentem USA je Barack Obama. Je tu ovšem nějaký pan Novák, který sice věděl, že kandidáty na úřad prezidenta jsou právě Barack Obama a John McCain, avšak těšně před vyhlášením výsledků voleb odjel na terénní výzkum do pralesů Jižní Ameriky, kde nemá žádný kontakt s civilizací. To znamená — za předpokladu, že se ještě nevrátil — že stále považuje za možné dva světy: svět, ve kterém je prezidentem USA Barack Obama, a svět, ve kterém je prezidentem USA John McCain (viz obr. 4.2, význam šípek si vysvětlíme později).



Obrázek 4.2: Možné světy

Shrnutí. Na příkladech výše jsme si demonstrovali aplikaci teorie možných světů v epistemické praxi. Konkrétněji, ukázali jsme si, jak je možné využít koncept možných světů k zachycení znalostí, resp. pochybností, agenta. Jak jsme si mohli dále všimnout, hlavní myšlenka v pozadí je ta, že agentův stav znalosti (epistemický stav) koresponduje s tím, do jaké míry je agent schopen „lokalizovat“ aktuální svět ve svém epistemickém univerzu, přičemž *epistemické univerzum* budeme chápat jako množinu

všech světů, které daný agent považuje za možné.

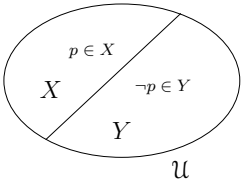
Agent samozřejmě ví, že aktuální (tj. ten pravdivý, skutečný) svět je ten svět, který obývá a v kterém započíná svoje epistemické úvahy, ale neví, který z množných světů to je. Ilustrovat si to můžeme na následujícím příkladu. Představme si, že jsme zamčeni ve svém hotelovém pokoji, ale tato skutečnost je nám utajena. Víme jen to, že náš pokoj má číslo 1408. Za pomoci předmětů v pokoji, výhledu z okna a dalších indicií se pak snažíme odvodit to, jestli je zvenku na dveřích hotelového pokoje skutečně číslo 1408. Podobně pak probíhá odhalování aktuálního světa mezi možnými světy.

Vedle teorie možných světů jsme se v příkladech výše seznámili i s několika dalšími klíčovými koncepty SEL. Prvním z nich bylo pojetí znalosti jako absence pochybnosti. V rámci SEL budeme brát znalost jako určitou pravdu obtisknutou do všech světů, které ten či onen agent považuje za možné, přičemž „mít znalost“ vlastně znamená „být schopen vyřadit část možných světů jako možných kandidátů na post aktuálního světa ze svého epistemického univerza“. Ty, které zůstanou, pak reprezentují *epistemické alternativy* daného agenta, tj. světy, které není schopen vyloučit a které pro něj představují potenciální aktuální světy. A samozřejmě, čím menší je ona množina možných světů, které agent považuje za možné, tím menší je jeho nejistota a tím více toho ví.

Posledními dvěma důležitými koncepty bylo chápání možných světů jako epistemických alternativ a s tím úzce související koncepce epistemické nerozlišitelnosti, jež mezi nimi panuje. Jedná se vlastně o určitou relaci, která spojuje epistemické alternativy, tj. světy, které od sebe agent není schopen rozlišit v tom smyslu, že neví, který z nich je pravdivý. Pokud si to shrneme, představili jsme si dohromady čtyři následu-

jící koncepce: (i) teorie možných světů, (ii) znalost jako absence pochybností, (iii) epistemické alternativy a (iv) epistemická nerozlišitelnost.

Všimněme si, že v pozadí možných světů je jednoduchý předpoklad, a to, že jakákoli agentova znalost rozděluje množinu všech možných světů na dvě podmnožiny, přičemž prvky první množiny představují ty světy, které jsou slučitelné s jeho znalostí, tj. jedná se o jeho epistemické alternativy, zatímco druhá množina je naopak tvořena těmi světy, které jsou v rozporu s jeho znalostí. Těm budeme říkat *kontra-epistemické* světy. Stručně řečeno, důsledkem připsání znalosti určitému agentovi je to, že se množina všech světů, které považuje za možné, rozštěpí na dvě části, tj. na epistemické alternativy a konstraepistemické světy, v závislosti na obsahu dané znalosti.



Obrázek 4.3: Množina možných světů

Pokud např. Alenka ví, že prší (Alenka ví, že p), pak se množina jejich možných světů rozdělí na dvě množiny: na nějakou množinu X, která bude obsahovat všechny možné světy kompatibilní s Alenčinou znalostí, že prší (tj. půjde o její epistemické alternativy), a na množinu Y obsahující ty světy, které jsou naopak s její znalostí neslučitelné (tj. půjde o kontraepistemické světy). Pod Y tedy budou spadat ty světy, ve kterých prší, je zamračeno, není slunečno atd. (viz obr. 4.3)

Z toho tedy vyplývá, že agent vylučuje ty světy, které jsou logicky

nekonzistentní s tím, co už ví. Proč například pan Novák z příkladů výše neuvažuje o možnosti, že oba dva kandidáti zahynuli spolu se zbytkem lidstva poté, co Zemi zasáhl obří meteorit? Tento katastrofický scénář může vyloučit, neboť kdyby tomu bylo skutečně tak, zahynul by i on. Avšak on žije, tudíž může tuto alternativu zavrhnout jako nemožnou. Analogicky, pokud pan Novák ví, že mu je 42 let, pak z množiny světů, které považuje za možné, může vyškrtnout všechny ty, ve kterých mu není 42 let, neboť jsou neslučitelné s tím, co ví.

Je potřeba si ale uvědomit, že agenti mají tendenci vylučovat i takové světy, které nejsou přímo v rozporu s jejich znalostmi. Když se Alenka procházela po deštivém Brně, neuvažovala o tom, že ji možná klame descartovský démon a že ve skutečnosti je slunečno. Stejně tak mohla Alenka v roli detektiva přijmout do svých úvah např. i takový svět, ve kterém se někdo snažil její bratry pouze falešně obvinit, ale neučinila tak. Podobně také pan Novák vynechal ze svých úvah svět, ve kterém byli oba dva kandidáti uneseni mimozemšťany, byť tuto možnost nic z jeho znalostí nevylučuje. Tím se dostáváme k dalšímu důležitému konceptu, a to k redukci možných světů.

4.1.4 Redukce možných světů

Redukce možných světů spočívá v odstranění určitých možných světů z epistemického univerza daného agenta.[10] Uvažme např. Descartův svět s klamavým démonem nebo virtuální světy jako tranCendenZ z Cronenbergova filmu *eXistenZ* (1999) či Matrix ze stejnojmenného snímku *Matrix* (1999) od sourozenců Wachowských. Tyto světy nejsou přímo

[10] Hendricks mluví o redukci jako o *stlačování* (angl. forcing). Srov. [Hendricks, 2006].

v logickém rozporu s našimi znalostmi, přesto je však do svých úvah běžně nezahrnujeme. Stejně tak ani běžně nezohledňujeme scénáře, ve kterých jsme jen mozkem v kádi[11] atd.

Principem redukce těchto možných světů je epistemická úspornost a ekonomičnost. Důvod je tedy spíše jen praktický. Naopak v případě pana Nováka došlo k redukci potenciálních možných světů na základě jeho přesvědčení o neexistenci mimozemských civilizací. V tomto případě byly principem redukce agentovy předsudky. Dalším principem redukce může být např. také náboženská víra, která zase ze sféry možných světů vyloučí vše, co je v rozporu s její naukou atd. Je tu tedy celá řada redukčních principů, které agent může využít k vymezení svého epistemického univerza. Když pak uvažuje o možných světech, přísně vzato neuvažuje skutečně o všech objektivně, logicky možných světech (démonské světy, mozky v kádi atd.), ale pouze o jejich určité podmnožině. Jinými slovy, některé z objektivně možných světů jsou neslučitelné s epistemologickým postojem agenta, a tudíž je *a priori* vyřazuje na základě odpovídajícího principu z množiny světů, které považuje za epistemicky možné. Jedná se vlastně o jisté předpoklady, které agent činí o svém epistemickém univerzu.[12] Pan Novák neuvažuje o všech možných světech, ale jen o těch, které považuje za možné on sám. Jeho epistemické univerzum je tedy podmnožinou všech objektivně možných světů.

Jen pro zdůraznění, zde nejde o štěpení epistemického univerza agen-

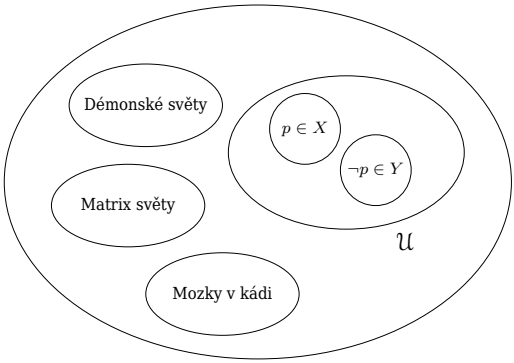
[11] Známý myšlenkový experiment Hilary Putnama, při kterém je náš mozek umístěn do kádě s umělou výživou a jeho nervová zakončení jsou zapojena do super počítače, který pomocí elektronických impulzů simuluje vnější realitu. Více viz [Putnam, 1981].

[12] Nutno poznamenat, že se nemusí ani vždy jednat o plně uvědomělý proces. Více viz např. Dretskeho teorie relevantních alternativ [Dretske, 1970] (angl. relevant alternatives theory).

ta na základě toho, co je slučitelné s tím, co ví nebo neví, tj. na epistemické alternativy a kontraepistemické světy, ale o principy redukce, na základě kterých se toto epistemické univerzum vůbec formuje. Rozklad na epistemické alternativy a kontraepistemické světy probíhá stále na základě logické konzistence, neboť předpoklad je takový, že pracujeme s racionálním agentem a takový agent může popírat existenci mimozemšťanů nebo klamavého démona bez pádných důkazů, ale rozhodně by neměl vědomě přijímat kontradikci.

Je třeba upozornit, že tato redukce probíhá mimo SEL v tom smyslu, že vymezuje doménu jejího působení. Vznikají tak vlastně dvě epistemická univerza, kterým můžeme říkat malé a velké (viz obr. 4.4 níže). Zatímco v rámci *malého epistemického univerza* (tj. v univerzu bez démonů, Matrixu apod.) žijeme naše každodenní životy (v rámci něj bude probíhat i SEL), *velké epistemické univerzum* přichází na řadu v podstatě jen v rámci filosofických debat. Konkrétní příklad: přestože Alenka ví, kde bydlí, tj. ve všech svých epistemických alternativách bydlí na stejném místě, rozhodně si dokáže představit, že by ji mohl klamat nějaký démon a ve skutečnosti bydlela úplně někde jinde. Jinými slovy, je ochotna připustit démonské světy jako teoreticky možné, ale nejsou pro ni pravými epistemickými alternativami. Považuje je sice za možné, ale nikoli možné ve smyslu potenciálně pravdivé, ale spíše možné jako logicky konzistentní. Je tedy třeba dávat pozor na to, kdy mluvíme o velkém epistemickém univerzu a kdy o malém epistemickém univerzu agenta, které je složené z epistemických alternativ a kontraepistemických světů. Pokud neuvedeme jinak, budeme epistemickým univerzem vždy mínit malé epistemické univerzum a značit jej budeme $\mathcal{U}$.

O malém epistemickém univerzu se někdy také mluví jako o rele-



Obrázek 4.4: Velké epistemické univerzum

vantních možných světech, tj. světech, o kterých má z pohledu agenta smysl uvažovat. Možné světy velkého epistemického univerza (démonské světy, Matrix světy, . . .) jsou vlastně vyloučeny z malého epistemického univerza jednoduše proto, že s ním nejsou kompatibilní, přesněji, nejsou slučitelné s epistemologickým postojem daného agenta. Předpokládejme, že je Alenka empirik. Pokud se podívá z okna a uvidí déšť, zformuje si přesvědčení „Vím, že prší“. Potom se pro ni představa, že venku ve skutečnosti neprší, že ji jen klame zlý démon, stává nekompatibilní (byť žádná její znalost tuto možnost nevylučuje), neboť je v rozporu s jejím epistemologickým stanoviskem. Abychom to shrnuli, SEL pracuje pouze v rámci malého epistemického univerza $\mathcal{U}$, přičemž možné světy velkého epistemického univerza, jsou světy, které agent vynechává ze svých úvah.

4.2 Výstavba SEL

Doposud jsme si SEL vymezili jen velmi stručně a neformálně jako modální intenzionální logiku s Kripkeho sémantikou možných světů, je to ale logický systém jako každý jiný, a skládá se tedy z jistých symbolů, syntaktických pravidel, axiomů, odvozovacích pravidel a z určité teorie modelů, resp. sémantiky.

Z technického hlediska půjde o systém jinak známý jako **S5**, který je díky svým vlastnostem „vlajkovou lodí" standardních epistemických logik. To ovšem neznamená, že by byl zcela neproblematický, jak si později ukážeme např. v souvislosti s tzv. pozitivní a negativní introspekcí. Přesto se však jedná o nejsilnější a v literatuře nejrozšířenější systém epistemické logiky (obzvláště mezi informatiky), a proto jej volíme jako výchozí bod.

Od klasické výrokové logiky se SEL ve své základní podobě liší ve dvou podstatných aspektech: (a) přibyl nový symbol K a axiomy a odvozovací pravidla definující chování tohoto symbolu a dále (b) modelově–teoretická sémantika je definována pomocí možných světů a relace epistemické dosažitelnosti. Na rozdíl od tradiční (aletické) modální logiky, která obsahuje operátor nutnosti $\Box$, tedy SEL pracuje s operátorem znalosti K. To znamená, že Kp (resp. $\Box p$) nebudeme číst jako „je nutné, že p", ale jako „je známo, že p".

Kripkeho sémantika zachovává platnost jak klasických tautologií výrokového kalkulu, tak i odvozovacího pravidla modus ponens (MP). V tomto ohledu je tedy SEL rozšířením výrokové logiky. Bylo sice řečeno, že výroková logika není vyhovující pro účely epistemické logiky, ale vzhledem k tomu, že nám nyní nejde o analýzu epistemických úsudků, ale

o výstavbu formálního modelu znalosti, zvolíme modální výrokovou logiku pro její jednoduchost.

V rámci dalšího zjednodušení budeme prozatím uvažovat jen o SEL obsahující jediného agenta, Alenku. Půjde tedy o *monoagentní systém*. Dále zanedbáme časový rozměr, takže náš model znalosti bude v tomto ohledu statický. Zatím nebudeme ani rozlišovat mezi epistemickou a doxastickou logikou v užším slova smyslu (o rozdílu mezi nimi se zmíníme až později). Tolik ke krátkému představení SEL a nyní se vrhneme na její formální vymezení.

4.2.1 Syntax SEL

Jak už jsme zmínili výše, v základu SEL leží jazyk výrokové logiky. Tento jazyk $\mathcal{L}$ nám bude sloužit k přesnému a jednoznačnému popisu epistemických situací, se kterými jsme se seznámili v příkladech výše (Alenka a počasí, Alenka řeší sudoku apod.). V důsledku tohoto omezení budeme tedy předpokládat, že náš agent uvažuje o světě (resp. o svém malém epistemickém univerzu) jen v rámci nějaké neprázdné množiny $\mathcal{P}$ atomických formulí, které budeme označovat symboly $p, q, r \ldots$ Tyto symboly budou zastupovat tvrzení o světě jako např. „prší v Brně“, „prší v Praze“, „Stanisław Lem je polský autor“ atd. Abychom mohli vyjádřit v SEL tvrzení jako „Alenka ví, že prší v Brně“, rozšíříme dále náš jazyk $\mathcal{L}$ o modální operátor pro znalost K. Formuli:

$$K_i p$$

budeme číst jako „agent i ví, že p“. Formule tvaru $K_i p$ tak vlastně zastupuje epistemické tvrzení, které vyjadřuje postoj agenta i k určitému

tvrzení p. A vzhledem k tomu, že SEL obsahuje jen jediného agenta, budeme index i vynechávat. Formuli jako:

$$Kp \wedge K(p \rightarrow q) \rightarrow Kq$$

pak můžeme číst jako „jestliže agent ví, že p a že p implikuje q, pak ví, že q“. Formule tvaru Kp lze číst také obecněji jako „je známo, že p“. Operátor K bude jediným modálním operátorem v našem systému SEL.[13]

Formálně vzato, jazyk $\mathcal{L}$ není nic jiného než jen množina formulí, přičemž začneme s atomickými formulemi v $\mathcal{P}$ a z nich utvoříme složitější formule pomocí negace $\neg$, konjunkce $\wedge$ a modálního operátoru K. Jazyk $\mathcal{L}$ je tedy formálně definován následovně:

Nechť $\mathcal{P}$ je neprázdná množina atomických formulí, pro jazyk $\mathcal{L}$ pak platí, že je to nejmenší množina taková, že:

- $\mathcal{P} \subseteq \mathcal{L}$,
- jestliže $\varphi \in \mathcal{L}$, pak $\neg\varphi \in \mathcal{L}$,
- jestliže $\varphi \in \mathcal{L}$ a $\psi \in \mathcal{L}$, pak $(\varphi \wedge \psi) \in \mathcal{L}$,
- jestliže $\varphi \in \mathcal{L}$, pak $K\varphi \in \mathcal{L}$,
- nic jiného není správně utvořená formule jazyka $\mathcal{L}$.

U formulí jako $(\varphi \wedge \psi)$ budeme vynechávat vnější závorky vždy, když to nepovede ke snížení srozumitelnosti. Také budeme používat standardní zkratky výrokové logiky jako $\varphi \vee \psi$ místo $\neg(\neg\varphi \wedge \neg\psi)$, $\varphi \rightarrow \psi$ místo $\neg\varphi \vee \psi$

[13] Jak si ale později ukážeme, je tu spousta dobrých důvodů, proč uvažovat i o systémech, které obsahují více než jen jednu modalitu.

a $\varphi \leftrightarrow \psi$ místo $(\varphi \rightarrow \psi) \wedge (\psi \rightarrow \varphi)$. Pravdivostní hodnotu pravda budeme zapisovat jako `true` a pravdivostní hodnotu nepravda jako `false`.

Syntax SEL musíme ještě rozšířit o axiomy a odvozovací pravidla. Prozatím nebudeme jejich volbu nijak komentovat, axiomy (resp. axiomatická schémata) a pravidla jen vyjmenujeme a vrátíme se k nim až později v sekci 4.3. Připomínáme, že se jedná o axiomy systému známého jako **S5**.

Axiomy

(P) Všechny tautologie výrokové logiky

(K) $K\varphi \wedge K(\varphi \rightarrow \psi) \rightarrow K\psi$ (axiom distribuce)

(T) $K\varphi \rightarrow \varphi$ (axiom pravdivosti)

(4) $K\varphi \rightarrow KK\varphi$ (axiom pozitivní introspekce)

(5) $\neg K\varphi \rightarrow K\neg K\varphi$ (axiom negativní introspekce)

Odvozovací pravidla

(MP) $\dfrac{\varphi \quad \varphi \rightarrow \psi}{\psi}$ (modus ponens)

(NEC) $\dfrac{\varphi}{K\varphi}$ (necesitace)

4.2.2 Sémantika SEL

Už jsme si popsali jazyk SEL, tj. množinu správně utvořených formulí, axiomy a odvozovací pravidla, budeme ovšem ještě potřebovat sémantiku, abychom byli schopni určit, kdy je nějaká formule φ pravdivá či

nepravdivá. Formule modální logiky se tradičně interpretují za pomoci možných světů, které formálně zachytíme v rámci Kripkeho sémantiky.

Předtím, než začneme, je ale nutné zmínit, že pojetí možných světů v rámci kripkovské sémantiky se poněkud liší od toho, které jsme si představili výše. Konkrétněji, možné světy zde nejsou chápany jako množiny atomických formulí, které v nich platí, ale jakožto atomické nestrukturované entity, na kterých nabývá každá atomická formule hodnoty pravda nebo nepravda. Tato dvě pojetí jsou pro naše účely ekvivalentní, nicméně je vhodné mít na paměti jejich rozdílnost (např. v kripkovském pojetí možných světů neplatí náš dřívější předpoklad, že není možné mít dva různé světy, ve kterých platí stejné atomické formule).

Kripkovská sémantika je tvořena Kripkeho *modely* $\mathcal{M}$, které reprezentují epistemický stav agenta. Kripkeho model $\mathcal{M}$, někdy rovněž nazývaný jako *struktura*, je trojice $\langle S, \mathcal{R}, \pi \rangle$ složená z neprázdné množiny S možných světů $S = \{s, t, u, \ldots, s_1, s_2, s_3, \ldots\}$, interpretace π a binární relace dosažitelnosti $\mathcal{R}$ na množině S. Dvojice $\langle S, \mathcal{R} \rangle$ se nazývá *rámec* modelu. Rámec můžeme reprezentovat jako orientovaný graf, kde uzly zastupují světy a hrany relaci dosažitelnosti $\mathcal{R}$. Interpretace π je funkce, která ve všech světech S daného modelu $\mathcal{M}$ přiřazuje každé atomické formuli p z $\mathcal{P}$ pravdivostní hodnotu. Model se tedy skládá z rámce a interpretační funkce.

Nechť $\mathcal{L}$ je jazyk výrokové modální logiky a $\mathcal{P}$ neprázdná množina atomických formulí, pak Kripkeho model $\mathcal{M}$ má tvar $\langle S, \mathcal{R}, \pi \rangle$, kde:

- S je neprázdná množina možných světů,
- $\mathcal{R}$ je binární relace epistemické dosažitelnosti $\mathcal{R} \subseteq S \times S$, tj. množina dvojic utvořená z prvků množiny S,

- $\pi(s) : \mathcal{P} \mapsto \{\texttt{true}, \texttt{false}\}$ je interpretační funkce, která přiřazuje pravdivostní hodnotu každé atomické formuli $p \in \mathcal{P}$ pro každý možný svět $s \in S$.

Jak už jsme si řekli, prvek S modelu $\mathcal{M}$ zastupuje množinu všech světů, které agent považuje za možné. Interpretace $\pi(s)$ nám slouží k tomu, abychom zjistili, zda je formule p pravdivá nebo nepravdivá v daném světě s. Jinými slovy, funkce π přiřazuje každé atomické formuli v každém světě pravdivostní hodnotu, resp. každé atomické formuli p přiřazuje množinu těch světů, ve kterých je p pravdivá. Jestliže tedy p označuje tvrzení „prší v Brně“, pak $\pi(s)(p) = \texttt{true}$ zachycuje ten stav věcí, kdy v Brně prší ve světě s. Interpretace je pak tradičním způsobem rozšířena na další formule prostřednictvím relace $\models$ (viz níže).

Úkolem binární relace epistemické dosažitelnosti $\mathcal{R}$ (či krátce jen relace $\mathcal{R}$) je spojovat ty světy, které jsou pro daného agenta epistemicky nerozlišitelné. Je třeba se uvědomit, že množina možných světů se může měnit a právě toto daná relace zachycuje. Jinými slovy, relace $\mathcal{R}$ vymezuje epistemické alternativy agenta. Skutečnost, že nějaké dva světy s a t jsou spojené relací $\mathcal{R}$ budeme zapisovat jako $(s, t) \in \mathcal{R}$,[14] což můžeme obecně číst jako „svět t je epistemicky možný ve světě s“. K relaci dosažitelnosti se ještě vrátíme v nadcházející sekci 4.2.2.1, prozatím jen dodáme, že po ní budeme vyžadovat jisté specifické vlastnosti.

Nyní definujeme, co to znamená, že určitá formule φ je pravdivá v daném světě s modelu $\mathcal{M}$, a to za pomoci binární relace $\models$ mezi světy a formulemi, které se říká *relace splnitelnosti* (angl. satisfaction relation). Všimněme si, že pravdivost formule φ bude záviset nejen na světě, ale

[14] Rovněž je možné se setkat se zkrácenými zápisy $\mathcal{R}(s, t)$ či $s\mathcal{R}t$.

i na modelu. Tuto skutečnost zapíšeme jako:

$$(\mathcal{M}, s) \models \varphi.$$

Čteme formule φ je pravdivá v $(\mathcal{M}, s)$, popř. φ je splnitelná (platná) ve světě s modelu $\mathcal{M}$. Formule jazyka $\mathcal{L}$ budou tedy interpretovány na dvojicích $(\mathcal{M}, s)$ složených z Kripkeho modelu $\langle S, \mathcal{R}, \pi \rangle$ a světů $s \in S$. Pár $(\mathcal{M}, s)$ budeme někdy zjednodušeně označovat jako (Kripkeho) svět.

Relaci $\models$ definujeme rekurzivně (induktivně) pomocí φ v $\mathcal{L}$ tak, že začneme s atomickými formulemi $p, q, r \ldots$ a postupně budeme pokračovat ke složitějším formulím φ, přičemž budeme předpokládat, že relace $\models$ byla definována pro všechny podformule φ. Onen člen π, který se vyskytuje v našem modelu $\mathcal{M}$, nám umožňuje vypořádat se se základním případem, kdy je φ atomická formule. Relace $\models$ tedy rozšiřuje π do všech formulí jazyka $\mathcal{L}$, a to podle následujících podmínek:

- $(\mathcal{M}, s) \models p \Leftrightarrow \pi(s)(p) = \texttt{true}$.

Řekneme, že atomická formule p je pravdivá, jestliže pro daný model $\mathcal{M}$ a svět $s \in S$ platí, že $(\mathcal{M}, s) \models p$. Konjunkce a negace se chová stejně jako ve výrokové logice, tj. konjunkce $\varphi \wedge \psi$ je pravdivá právě tehdy, když obě formule φ a ψ jsou pravdivé, a negovaná formule $\neg\varphi$ je pravdivá právě tehdy, když φ je nepravdivá. Formálněji:

- $(\mathcal{M}, s) \models \neg\varphi \Leftrightarrow (\mathcal{M}, s) \not\models \varphi$,
- $(\mathcal{M}, s) \models \varphi \wedge \psi \Leftrightarrow (\mathcal{M}, s) \models \varphi$ a $(\mathcal{M}, s) \models \psi$.

Všimněme si, že podmínka pro negaci zaručuje, že naše logika bude dvouhodnotová. To znamená, že pro každou formuli φ budeme mít buď

$(\mathcal{M}, s) \models \varphi$, nebo $(\mathcal{M}, s) \models \neg\varphi$, ale nikdy ne obojí. Nechť $\mathbb{M}$ je množina všech Kripkeho modelů $\mathcal{M}$, pak $\mathbb{M} \models \varphi$ znamená, že pro všechna $\mathcal{M} \in \mathbb{M}$ platí, že $\mathcal{M} \models \varphi$.

Nyní si představíme pojmy platnosti a splnitelnosti. V rámci Kripkeho modelu $\mathcal{M} = \langle S, \mathcal{R}, \pi \rangle$ řekneme, že formule φ je *platná* v $\mathcal{M}$ a zapíšeme $\mathcal{M} \models \varphi$, jestliže $(\mathcal{M}, s) \models \varphi$ pro všechny světy $s \in S$. A řekneme, že formule φ je *splnitelná* v $\mathcal{M}$, jestliže $(\mathcal{M}, s) \models \varphi$ pro některý svět $s \in S$. Dále řekneme, že φ je *platná* a napíšeme $\models \varphi$, jestliže je φ platná v množině všech modelů $\mathbb{M}$, a že φ je splnitelná, jestliže je φ splnitelná v nějakém modelu $\mathcal{M} \in \mathbb{M}$. Bude-li to umožňovat kontext, budeme zápis $\models \varphi$ zkracovat jen na φ. Formule φ je platná v modelu $\mathcal{M}$ právě tehdy, když formule $\neg\varphi$ není v tomto modelu splnitelná. A naopak, φ je splnitelná právě tehdy, když formule $\neg\varphi$ není v daném modelu $\mathcal{M}$ platná.

Už tedy víme, jak lze v SEL zapsat, že nějaká formule φ je pravdivá. Jak ale zapsat, že někdo ví, že tato formule je pravdivá? Co to vlastně znamená, když o agentovi řekneme, že ví, že nějaké φ je pravdivé? Tím se konečně dostáváme k formulím tvaru $K\varphi$. Nyní se pokusíme formálně zachytit intuici, kterou jsme si představili v příkladech s Alenkou výše, tj. že agent ví, že φ právě tehdy, když φ je pravdivé ve všech světech, které agent považuje za možné. Formálně:

- $(\mathcal{M}, s) \models K\varphi \Leftrightarrow (\mathcal{M}, s) \models \varphi$ pro všechna t taková, že $(s, t) \in \mathcal{R}$.

Čteme: (v modelu $\mathcal{M}$ a světě s) agent ví, že φ právě tehdy, když φ je pravdivé ve všech světech t, tj. ve všech epistemických alternativách, které agent považuje ve světě s za možné. Nebo obráceně, jestliže je formule φ pravdivá ve všech světech, které agent považuje za možné, pak řekneme, že tento agent ví, že φ. To znamená, že o agentovi řekneme, že

neví nějaké φ právě tehdy, když φ je nepravdivé alespoň v jednom ze světů, které považuje za možné. Pokud si to shrneme, pravdivostní hodnota formulí, které neobsahují symbol K, závisí pouze na pravdivostní hodnotě atomických formulí daného světa. Jde-li o formule tvaru $K\varphi$, jejich pravdivostní hodnota ve světě s závisí na pravdivostní hodnotě vnořené formule φ ve světech dosažitelných z daného světa s.

Uvažme např. formuli $(\neg p \wedge \neg q) \wedge (K\neg p \wedge \neg K\neg q)$, která je v SEL splnitelná pro nějaké $\mathcal{M}, s$, tj. $(\mathcal{M}, s) \models (\neg p \wedge \neg q) \wedge (K\neg p \wedge \neg K\neg q)$. To znamená, že přestože mají formule p a q stejnou pravdivostní hodnotu, agent může znát negaci jedné, aniž by znal druhou. Jedná se vlastně formální vyjádření situace, kterou známe už z dřívějška. Vzpomeňme si na náš případ s Alenkou, která přemýšlela o počasí. Předpokládejme, že v Brně i v Praze je slunečno, tj. formule $\neg p \wedge \neg q$ je pravdivá, avšak zatímco Alenka ví, že v Brně je slunečno, resp. že neprší $K\neg p$, o Praze to neví, $\neg K\neg q$. Pokud to dáme celé dohromady, získáme právě $(\mathcal{M}, s) \models (\neg p \wedge \neg q) \wedge (K\neg p \wedge \neg K\neg q)$.

4.2.2.1 Relace epistemické dosažitelnosti

Pokud jde o relaci $\mathcal{R}$, řekli jsme si, že spojuje epistemické alternativy. S tímto termínem jsme se už setkali, nyní jen o něco zpřesníme jeho význam. Jako epistemické alternativy budeme chápat ty světy, které agent není schopen rozlišit od aktuálního světa. Množina všech epistemických alternativ tak vlastně zprostředkovaně představuje agentovu neznalost a nejistotu o tom, jaký je svět doopravdy, a tím pádem vlastně i jeho znalost.

Představme si situaci, modelovanou jako možný svět s, který je zároveň i světem aktuálním, kdy Alenka uvažuje o tom, jestli má v led-

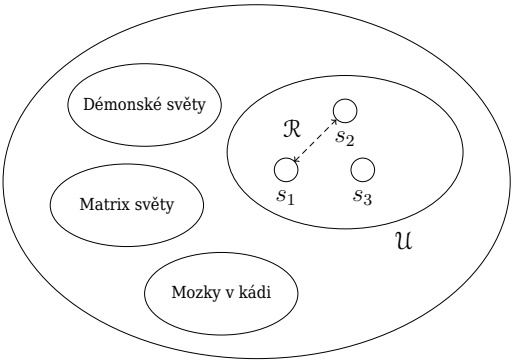
nici mléko. Jsou v podstatě jen dvě možnosti (dva možné světy), buď tam mléko je (svět s), nebo tam mléko není (svět t). Řekli jsme si, že svět s budeme považovat za aktuální, tedy mléko v lednici skutečně je, nicméně to Alenka neví. S informacemi, které má k dispozici, jsou pro ni oba dva světy s a t stejně možné. Jinými slovy, Alenka neví, jestli je v lednici mléko, tj. neví, který ze světů s a t je ten aktuální. A právě tuto situaci zapíšeme jako $(s, t) \in \mathcal{R}$.[15]

Co je to tedy binární relace epistemické dosažitelnosti? Je to relace, která spojuje rovnocenné, tj. epistemicky nerozlišitelné světy. Stručně řečeno, dva možné světy jsou propojené relací $\mathcal{R}$ právě tehdy, když v nich má agent stejné znalosti. Relace $\mathcal{R}$ tak vlastně propojuje výchozí (aktuální) svět se všemi jeho alternativami, které jsou pro daného agenta epistemicky nerozlišitelné. Je třeba opět připomenout, že agent neví, který svět je ten aktuální, tj. i aktuální svět je pro něj jen jedním z možných světů. Kdyby agent věděl, který svět je aktuální, znamenalo by to, že množina světů, které považuje za možné, obsahuje jen jeden svět, a tedy že má nulovou nejistotu. Jinými slovy, agent by disponoval absolutní znalostí všeho.

Vzpomeňme si na situaci, kdy se Alenka rozhodovala o tom, komu připsat autorství dramatu *R.U.R.* Obě možnosti, reprezentovány v obr. 4.5 světy s_1 a s_2, pro ni byly ekvivalentní, tj. neměla důvod jednu epistemickou alternativu upřednostnit před druhou, což je v obrázku naznačeno relací $\mathcal{R}$, zatímco svět s_3 znázorňuje vyloučenou možnost „Autorem *R.U.R.* je Isaac Asimov“. Je to tedy právě relace $\mathcal{R}$, která rozděluje malé epistemické univerzum na epistemické alternativy a kontraepistemické

[15] Jak správně upozorňuje [Holliday, 2013], zápis $(s, t) \in \mathcal{R}$ bychom tedy neměli přísně vzato interpretovat jako „Alenka považuje ve světě s svět t za možný“, neboť toto čtení může vzbuzovat dojem jakési psychologické spíše než epistemické možnosti.

světy. Nyní blíže k relaci $\mathcal{R}$.



Obrázek 4.5: Ekvivalentní světy

Je zřejmé, že množina epistemických alternativ, které agent považuje za možné, se přímo odvíjí od jeho aktuálního stavu informací neboli jeho *báze znalostí* (angl. knowledge base). A je to právě tato závislost, kterou formálně zachycuje binární relace $\mathcal{R}$ na množině možných světů. Jinými slovy, relace $\mathcal{R}$ zachycuje myšlenku, že agent považuje ve světě s svět t za (epistemicky) možný, pokud jsou stavy věcí (resp. pravdivostní ohodnocení) světa t slučitelné s jeho bází znalostí, tj. s tím, co ví v aktuálním světě s.

Už víme, že relace $\mathcal{R}$ spojuje ty světy, které agent nedokáže rozlišit, přesněji, ty, které jsou pro něj z epistemického hlediska rovnocenné neboli ekvivalentní. To znamená, že budeme s relací $\mathcal{R}$ v rámci systému SEL (resp. **S5**) nakládat jako s *relací ekvivalence*. Relace ekvivalence je binární relace, která splňuje následující tři podmínky:

- reflexivita, tj. $(s, s) \in \mathcal{R}$ pro všechna $s \in S$,
- symetrie, tj. $(s, t) \in \mathcal{R} \Leftrightarrow (t, s) \in \mathcal{R}$ pro všechna $s, t \in S$,

- tranzitivita, tj. $\big((s,t) \in \mathcal{R} \wedge (t,u) \in \mathcal{R}\big) \Rightarrow (s,u) \in \mathcal{R}$ pro všechna $s,t,u \in S$.

Všimněme si, že podmínka reflexivity na relaci $\mathcal{R}$ zachycuje náš předpoklad, že pravdivost je nutnou podmínkou znalosti. Přesněji řečeno, reflexivita zaručuje, že aktuální svět (tj. ten pravdivý, skutečný svět) bude vždy agentem považován za jeden z možných světů. A vzhledem k tomu, že jsme definovali znalost jako něco, co musí být pravda ve všech dosažitelných světech, je jasné, že dokud bude aktuální svět součástí agentových úvah (byť maskovaný jako jeden z možných světů), nikdy nemůže vědět něco, co není pravda.[16]

Symetrie zaručuje to, že pokud bude svět t přístupný ze světa s, pak bude také svět s zpětně dosažitelný ze světa t, tj. umožňuje obousměrnou dosažitelnost. Na první pohled se to může jevit zcela samozřejmé, ale jak si později ukážeme, nemusí tomu tak být vždy: např. v některých z možných světů t (kdyby byly skutečně aktuální) by mohl mít agent k dispozici jiné, odlišné či dodatečné informace, které by vyloučily svět s jako možný, tj. agent jakoby nedomyslí všechny důsledky, popř. předpoklady určitého možného světa do konce. Obrazně řečeno, může se ukázat, že cesta ze světa s do světa t je jednosměrná.

Tranzitivita označuje tu vlastnost, že pokud je nějaký svět t dosažitelný se světa s a nějaký svět u je dosažitelný ze světa t, pak je také svět u dosažitelný ze světa s.

To, že po relaci $\mathcal{R}$ vyžadujeme, aby se chovala jako relace ekvivalence, se zdá být rozumným předpokladem, neboť nám umožňuje zachytit to,

[16] Na příkladech níže si ovšem ukážeme, že situace nemusí být vždy tak jednoduchá. Můžeme si totiž představit např. takového agenta, který nepovažuje skutečný svět za možný.

že agent považuje ve světě s svět t za možný, jestliže v obou světech s a t má stejné informace, tj. když jsou pro něj světy s a t epistemickými alternativami.

Je potřeba ale zmínit, že relace ekvivalence (resp. systém **S5**) zachytává poměrně silný pojem znalosti, který nemusí být vždy přímo žádoucí. Slabších systémů (a tedy i pojmů znalosti) lze dosáhnout jinou volbou vlastností relace $\mathcal{R}$. Tato relace může být např. reflexivní, tranzitivní, ale ne symetrická. To je nesmírně důležité, neboť vlastnosti celého modelu znalosti se odvíjí právě od toho, jaké vlastnosti relaci $\mathcal{R}$ přiřkneme. Uvažme následující situace, ve kterých relace $\mathcal{R}$ není relací ekvivalence.

Předpokládejme, že bůh neexistuje. Rozhodně si můžeme představit, že náš agent je věřící, a tedy takovou možnost nepřipouští, byť informace, které má přímo k dispozici, tuto možnost nevylučují (není dokázáno, že bůh existuje). Jinými slovy, ve všech světech, které náš agent považuje za možné, existuje bůh. Nicméně náš předpoklad byl takový, že bůh v aktuálním světě neexistuje. A vzhledem k tomu, že i aktuální svět je jedním z možných světů, výsledek je takový, že agent nepovažuje za možné, že by bůh neexistoval, i když tomu tak (za našeho předpokladu) je. Toto je případ, ve kterém relace $\mathcal{R}$ ztrácí reflexivitu. Agent nebyl schopen zahrnout do svých úvah jeden z možných světů, který byl shodou okolností světem aktuálním.

Uvažme další příklad: mějme svět s, ve kterém Fredova manželka Harriet jela navštívit svoji přítelkyni Alici, přičemž tuto skutečnost Fredovi s předstihem sdělila. Fred na to ovšem zapomněl, takže považuje za možný i svět t, ve kterém Harriet jela navštívit svého bratra Roberta. Ale kdyby Harriet řekla, že jede navštívit Roberta, Fred by tuto informaci nezapomněl, jelikož před týdnem měla Harriet s Robertem ošklivou hádku.

Tedy ve světě t by Fred svět s nepovažoval za možný, jelikož ve světě t by si Fred pamatoval, že jeho manželka jela navštívit Roberta. Je sice možné, že by si Fred po nějakém hlubším reflektování svých znalostí uvědomil, že svět t není možný, protože v t by si pamatoval, co Harriet řekla, nicméně ne všichni a vždy podrobují své znalosti tak důkladnému přezkoumání. Toto je případ, ve kterém relace $\mathcal{R}$ ztrácí symetrii. Agent vzal do svých úvah takový svět t, ze kterého není jeho aktuální svět s zpětně dosažitelný.[17]

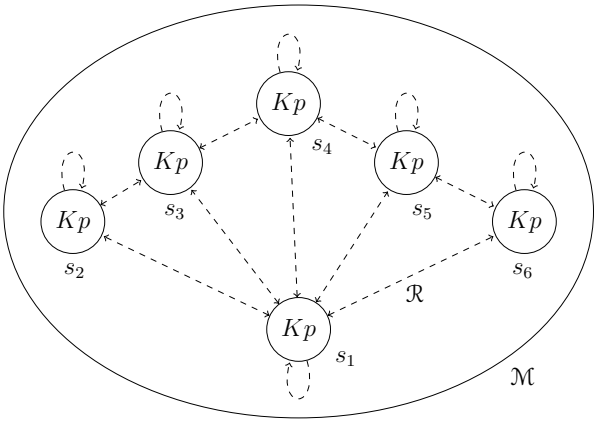
Tyto a jim podobné případy je třeba mít stále na paměti, jelikož odlišné vlastnosti relace $\mathcal{R}$ odpovídají odlišnému pojetí znalosti. Přestože volba relace $\mathcal{R}$ jako relace ekvivalence se zdá být dobře zdůvodněná, v mnoha případech se může tento předpoklad na relaci $\mathcal{R}$ ukázat jako nepatřičný (viz příklady výše), nesmíme tedy zapomínat na to, že je to jen jedna z více možností.

Jak už bylo zmíněno, jednou z výhod Kripkeho sémantiky, resp. modelů $\mathcal{M}$, je to, že si je můžeme znázornit pomocí jednoduchých orientovaných grafů,[18] kde vrcholy budou znázorňovat možné světy a hrany grafu budou zastupovat relaci $\mathcal{R}$. To znamená, že dva možné světy (vrcholy) budou propojené relací $\mathcal{R}$ (hranou) právě tehdy, jestliže tyto dva možné světy jsou pro agenta epistemicky nerozlišitelné (za předpokladu, že $\mathcal{R}$ je relací ekvivalence). Předpokládejme, že agent uvažuje o šesti možných světech s_1 až s_6 (viz obr. 4.6).

Situaci na obr. 4.6 můžeme popsat takto: agent nezná všechny fakty vnějšího světa (tj. zkrátka nějaké věci nezná), takže považuje více světů

[17] Tento příklad byl vypůjčen z [Fagin et al., 1995].

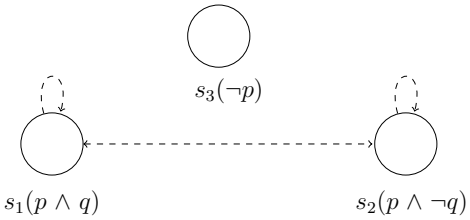
[18] Orientovaný graf je dvojice $G = \langle V, E \rangle$, kde V je neprázdná množina *vrcholů* (uzlů) a $E \subseteq V \times V$ je množina uspořádaných dvojic vrcholů, tj. (orientovaných) *hran*.



Obrázek 4.6: Epistemická nerozlišitelnost

za možné. V našem případě uvažuje konkrétně o šesti možných světech $s_1 - s_6$, tj. pro všechny světy $s_2 - s_6$ platí, že $(s_1, s_n) \in \mathcal{R}$, přičemž s_1 budeme pokládat za aktuální svět. Tento graf ovšem není moc užitečný, neboť zachycuje jen to, co agent ví, v našem případě p. Obecně je samozřejmě lepší do grafu zanášet opačné údaje, tj. ty informace, kterými se od sebe jednotlivé světy liší. Graf výše také zachycuje předpoklady, které jsme učinili o relaci $\mathcal{R}$. Smyčky u jednotlivých světů jsou důsledkem toho, že relace $\mathcal{R}$ je reflexivní, tj. každý svět je dosažitelný ze sebe sama, šipky na obou koncích hran znázorňují to, že $\mathcal{R}$ je symetrická, a snadno si můžeme ověřit i to, že platí tranzitivita.

Zkusme ještě jeden, jednoduší příklad. Co můžeme z grafu na obr. 4.7 vyčíst Předpokládejme, že s_1 představuje aktuální svět, tj. že skutečný stav věcí je takový, že je pravda, že $p \wedge q$. Agent tedy považuje dva světy za možné: aktuální svět s_1 a potenciální svět s_2. A vzhledem k tomu, že q je v jednom ze světů pravdivé a v druhém nepravdivé, můžeme (s vy-



Obrázek 4.7: Epistemická alternativa

užitím naší definice znalosti) odvodit, že daný agent neví, zda platí q. Jinak řečeno, není se schopen rozhodnout mezi tím, zda je pravda q či $\neg q$. Analogicky pak můžeme odvodit, že ví, že platí p, neboť p je pravdivé ve všech světech, které považuje za možné, tj. s_1 a s_2. Svět s_3 je do grafu zanesen z čistě ilustračních důvodů a znázorňuje to, že agent nepovažuje svět s_3 za možný, a stává se tak pro něj kontraepistemickým světem.

4.3 SEL: bližší pohled

V předcházející části jsme definovali syntax a sémantiku SEL. Nyní blíže prozkoumáme axiomy a odvozovací pravidla, kterými se tento model znalosti vyznačuje. Jen pro připomenutí, relaci $\mathcal{R}$ budeme pojímat jako relaci ekvivalence.

Jednou ze základních vlastností našeho modelu znalosti SEL je to, že agentova znalost je uzavřena vzhledem k materiální implikaci. Jestliže agent ví, že φ a že $\varphi \rightarrow \psi$, pak z naší definice znalosti výše vyplývá, že jak φ, tak $\varphi \rightarrow \psi$ musí být pravdivé ve všech světech, které považuje za možné. Tedy i ψ musí být pravdivé ve všech světech, které agent pova-

žuje za možné, v důsledku čehož agent musí vědět ψ. Formálně:

(K) $K\varphi \wedge K(\varphi \to \psi) \to K\psi$

Tato formule se nazývá *axiom distribuce*, jelikož nám doslova umožňuje „distribuovat" operátor K přes implikaci.[19] Axiom **(K)** je někdy také nazýván jako K–axiom, axiom logické racionality či axiom logické vševědoucnosti. Jak si můžeme všimnout, nejedná se o nic jiného než o epistemický uzávěr, tj. uzavřenost znalosti vzhledem k implikaci, problém dobře známý i tradiční epistemologii.[20] Tento axiom nám vlastně říká, že pokud agent zná obě premisy odvozovacího pravidla modus ponens, pak je v našem systému SEL zavázán k tomu, aby věděl i závěr.[21]

Přestože se tento axiom může zdát jako nanejvýše rozumný, nelze mu přiřknout všeobecnou platnost. Připomeňme si úsudek (eMP):

$$\text{(eMP)}\ \frac{\begin{array}{c}\text{Alenka ví, že prší.}\\ \text{Alenka ví, že jestliže prší, pak není pravda, že neprší.}\end{array}}{\text{Alenka ví, že není pravda, že neprší.}}$$

Alenka může znát obě premisy, aniž by znala závěr, konkrétně že není pravda, že neprší. Tuto možnost ovšem systém SEL nepřipouští. Obecně, pokud agent ví, že platí jak φ, tak i $\varphi \to \psi$, pak nemůže nevědět, že platí i ψ.

Uvažme další vlastnost: jestliže je nějaká formule φ platná ve všech objektivně (logicky) možných světech (tj. je-li logickou pravdou), pak φ musí být rovněž pravdivá i ve všech světech, které agent považuje

[19] Axiom **(K)** se někdy také uvádí v alternativní podobě $K(\varphi \to \psi) \to (K\varphi \to K\psi)$.

[20] Srov. např. [Dretske, 1970], [Nozick, 1981].

[21] Jak už jsme dříve zmiňovali, **(K)** je vlastně epistemická varianta pravidla modus ponens formulovaná jako axiom.

za možné, neboť jsou podmnožinou všech objektivně možných světů. Z toho vyplývá, že formule $K\varphi$ je pravdivá ve všech epistemických alternativách daného agenta vždy, když je formule φ objektivně platná, tj. tautologií (teorémem). Formálněji:

(NEC) $\dfrac{\varphi}{K\varphi}$

Slovy: z dokázaného φ odvoď $K\varphi$.[22] Tato vlastnost se nazývá *pravidlo generalizace znalosti*, popř. epistemická necesitace, a vystupuje jako jedno z odvozovacích pravidel. Dozvídáme se z něj, že jestliže je formule φ teorémem, pak ji také známe. Jinak řečeno, pokud je formule φ platná, pak je platná i formule $K\varphi$. Toto pravidlo nesmíme zaměnit s formulí $\varphi \to K\varphi$, která říká, že pokud je formule φ pravdivá, pak ji agent zná. Agent nezná nutně všechna tvrzení, která jsou pravdivá, nicméně — a tomu se v našem modelu nijak nevyhneme — agent zná všechny platné formule jako např. „V Praze teď prší, nebo v Praze teď neprší“ apod. Stručně řečeno, pokud je nějaká formule logickou pravdou, pak ji agent zná.

Je třeba zdůraznit, že jak **(K)**, tak i **(NEC)** nevyžadovaly žádné předpoklady ohledně relace $\mathcal{R}$, a jsou tedy platné ve všech Kripkeho modelech. Jedná se vlastně o charakteristické vlastnosti Kripkeho sémantiky možných světů. Modálním logikám, ve kterých platí **(K)**, **(NEC)** a **(MP)**, se říká *normální* modální logiky.

Přestože agent nemusí znát všechna pravdivá tvrzení, pokud nějaká zná, pak musí být pravdivá. Formálně:

(T) $K\varphi \to \varphi$

[22] Všimněme si, že **(NEC)** je zkrácený zápis. Nezkrácená verze pravidla by vypadala „je-li φ teorém, pak lze odvodit další teorém $K\varphi$“.

Axiom **(T)** zachycuje jednu z našich nejzákladnějších intuic ohledně znalosti, a to že nemůžeme vědět něco, co není pravda. Pokud „víme“ něco, co není pravda, pak to zkrátka nevíme. Tedy pokud Alenka řekne např. „Vím, že $p \supset q$ je teorém“, buď lže, nebo to prostě ve skutečnosti neví a jen si myslí (věří, doufá), že to ví. Je to právě tato vlastnost, někdy nazývaná také jako *axiom znalosti* či *axiom pravdivosti*, která od sebe odděluje logiku znalosti (tj. epistemickou logiku v užším smyslu) a logiku přesvědčení (tj. doxastickou logiku).

Je zřejmé, že u přesvědčení nemá smysl vyžadovat stejnou závaznost vůči pravdivosti jako v případě znalosti. Jestliže je Alenka přesvědčena o tom, že existuje Ježíšek, nevyplývá z toho, že Ježíšek musí existovat. Ovšem pokud by to Alenka věděla, pak bychom z toho mohli odvodit, že Ježíšek skutečně existuje. Krátce řečeno, věřit můžeme úplně všemu, ale vědět můžeme jedině pravdu.

Poslední dvě vlastnosti, o kterých se v této části zmíníme, umožňují agentovi jistou formu introspekce či reflexe vlastních znalostí. Zaprvé, agent ví to, co ví, formálně:

(4) $K\varphi \to KK\varphi$

Axiom **(4)**, rovněž známý jako *axiom pozitivní introspekce* či KK-axiom, zaručuje, že si agent bude vědom toho, co sám ví, tj. dokáže reflektovat svoje vlastní znalosti. Jedná se o problematický předpoklad, který nemusí zdaleka vždy platit. Jako tradiční protipříklad se uvádí tzv. *nevědomá znalost* (angl. unconscious knowledge). Představme si následující situaci: Alenka ráno vstane, zapne rádio a pustí se do snídaně. Během toho, co snídá, zazní v rádiu informace „největším známým člověkem byl Robert Wadlow měřící 2,72 metrů“, kterou však Alenka přeslechne,

neboť byla naplno zabraná do jídla. Poté co dosnídá, vyrazí do školy, kde na ni čeká neohlášený test ze všeobecného přehledu. Jedna z otázek zní: „Jak se jmenoval nejvyšší člověk na světě měřící 2,72 metrů?“, přičemž na výběr má ze šesti možností, z nichž jedna je právě Robert Wadlow. Alenka netuší správnou odpověď, ale tipne si, že je to Robert Wadlow. Její tip ovšem nebyl tak úplně tipem a možnost „Robert Wadlow“ jí padla do oka právě proto, že to dříve zaslechla. Jinými slovy, přestože Alenka tuto informaci ráno vědomě nezpracovala, její mozek ji zaregistroval. Alenka tedy tipuje jen domněle a ve skutečnosti, byť podvědomě, to ví (resp. její mozek to ví).[23]

Zadruhé, agent dále ví i to, co neví, formálně:

(5) $\neg K\varphi \rightarrow K\neg K\varphi$

Axiom **(5)** neboli *axiom negativní introspekce*, někdy také nazýván jako axiom moudrosti, platónský axiom či sokratovský axiom, zajišťuje to, aby si agent byl vědom toho, co neví. Jak název napovídá, jedná se vlastně o opačnou verzi pozitivní introspekce. Zatímco ta umožňovala agentovi reflektovat svoje vlastní znalosti, negativní introspekce mu umožňuje uvědomovat si svoji vlastní neznalost. Jinými slovy, pokud agent něco neví, pak ví, že to neví.

Je zřejmé, že tato podmínka je ještě náročnější a problematičtější než ta kladená na agenta pozitivní introspekcí. Proč je tomu tak? Vypomůžeme si jednoduchým příkladem. Předpokládejme, že Alenka neví, že existuje něco takového jako ptakopysk. Nikdy o nich neslyšela, nikdy je neviděla, a dokonce ani nikdy neslyšela, ani neviděla výraz „ptakopysk“.

[23] Podobných protipříkladů je celá řada. Zmínit můžeme např. obličejovou slepotu (prosopagnosie), kdy pacient nepoznává obličeje svých blízkých, přestože jeho emocionální odezva (odvozená např. ze zvýšené vodivosti kůže) vypovídá o opaku.

To, že Alenka neví, co je to ptakopysk, zapíšeme jako $\neg Kp$. Ovšem podle axiomu negativní introspekce by z toho Alenka mohla vyvodit, že si je vědoma (že ví), že neví, co je to ptakopysk, aniž by výraz ptakopysk znala. To je zcela nepřijatelný závěr. Agent totiž nemusí být vůbec obeznámen s výrazem, o který se jedná. Pokud by tvrzení „Nevím, že existuje něco takového jako ptakopysk" pronesla samotná Alenka, tak je samozřejmě zcela rozumné předpokládat, že by z toho dokázala odvodit i to, že ví, že to neví. Nicméně jistě jsou tu takové oblasti vědění, o kterých nemá Alenka ani tušení, že existují, natož aby věděla, že je nezná.

Uvažme podobný příklad: pokud na ulici přistoupíme k nějaké osobě a zeptáme se jí: „Víte, kdo je to Flash Gordon?", může samozřejmě odpověď „Ano" nebo „Ne" v závislosti na svých znalostech. Nicméně pokud toto jméno slyšela dotyčná osoba poprvé v životě, je velmi pochybné předpokládat, že měla v paměti (resp. znalostní bázi) uloženou informaci „Vím, že nevím, kdo je to Flash Gordon" ještě předtím, než vůbec zjistila, že existuje nějaký Flash Gordon. Ovšem přesně k tomu nás axiom negativní introspekce zavazuje. O agentech, kteří jsou schopni jak pozitivní, tak i negativní introspekce, se někdy mluví jako o plně introspektivních agentech.

Shrnutí. Dohromady jsme si představili pětici základních vlastností našeho modelu znalosti. Jak už jsme si řekli, nejedná se v podstatě o nic jiného než o axiomy a odvozovací pravidla, které definují chování našeho unárního epistemického operátoru K. Pro všechny formule φ a ψ a všechny modely $\mathcal{M}$, kde relace $\mathcal{R}$ je relace ekvivalence, tedy platí **(K)**, **(NEC)**, **(T)**, **(4)** a **(5)**.

Skupina těchto pěti vlastností, jmenovitě, definuje systém známý jako

S5. Systém **S5** je korektní a úplný vzhledem k danému axiomatickému systému, tj. všechny axiomy jsou platné a všechny platné formule mohou být odvozeny z daných axiomů **(K)** až **(5)**.[24]

Připomeňme si, že platnost těchto modálních axiomů se přímo odvíjí od vlastností jako reflexivita, symetričnost atd., které přiřkneme relaci $\mathcal{R}$. V systému **S5** jsou vedle **(NEC)** platná ještě dvě následující pravidla odvození:

$$\textbf{(MON)} \quad \frac{\varphi \to \psi}{K\varphi \to K\psi}$$

$$\textbf{(CON)} \quad \frac{\varphi \leftrightarrow \psi}{K\varphi \leftrightarrow K\psi}$$

která jsou důsledkem **(K)** a **(NEC)**. Pravidlo **(MON)** neboli monotonie v podstatě říká, že pokud je $\varphi \to \psi$ platnou formulí, pak $K\varphi \to K\psi$ je rovněž platnou formulí. Jinými slovy, z $\varphi \to \psi$ můžeme odvodit $K\varphi \to K\psi$. Pravidlo kongruence **(CON)** nám říká totéž, ovšem v případě materiální ekvivalence. Platnost **(K)** a **(NEC)** (a tedy i **(MON)** a **(CON)**) zajišťuje samotný Kripkeho model, platnost **(T)** vyplývá z toho, že relace $\mathcal{R}$ je reflexivní, platnost **(4)** je důsledkem toho, že $\mathcal{R}$ je tranzitivní a platnost **(5)** je zaručena tím, že $\mathcal{R}$ je relací ekvivalence, tj. je reflexivní, symetrická a tranzitivní (resp. reflexivní a eukleidovská).

Je však systém vymezený právě touto pětící vlastností adekvátním modelem znalosti? Jak už jsme si řekli dříve, přestože se rozhodnutí považovat relaci $\mathcal{R}$ zrovna za relaci ekvivalence jeví jako rozumný krok z mnoha hledisek, jsou tu i další možnosti. Vzpomeňme si třeba na případ s Fredem a jeho manželkou Harriet. Modifikací relace $\mathcal{R}$ můžeme

[24] Srov. [Chellas, 1980].

získat nové modely znalosti s odlišnými vlastnostmi. Těmito systémy se budeme zabývat nyní.

4.3.1 Od K až k S5: cesta tam a zase zpátky

V této části si představíme blíže některé z nejběžnějších modelů znalosti, které můžeme získat různými restrikcemi na relaci $\mathcal{R}$. Začneme u základního a nejslabšího modelu znalosti **K**. Jazyk $\mathcal{L}$ bude opět množina formulí, které mohou být sestaveny z atomických formulí v $\mathcal{P}$ za použití konjunkce, negace a modálního operátoru K. O relaci $\mathcal{R}$ doposud nebyly učiněny žádné předpoklady. Model znalosti **K** je definován následujícími axiomy a odvozovacími pravidly:

Axiomy

(P) Všechny tautologie výrokové logiky

(K) $K\varphi \wedge K(\varphi \rightarrow \psi) \rightarrow K\psi$

Odvozovací pravidla

(MP) $\dfrac{\varphi \quad \varphi \rightarrow \psi}{\psi}$

(NEC) $\dfrac{\varphi}{K\varphi}$

Sémantika je definována stejným způsobem jako v **S5**. Systém **K** je korektní (tj. je-li φ teorémem, pak je φ i tautologií) a úplný (tj. je-li φ tautologií, pak je φ i teorémem) vzhledem k množině všech Kripkeho modelů. **K** je tedy minimální epistemickou logikou, která obsahuje všechny platné

formule výrokové logiky společně s axiomem **(K)**, a je uzavřena vzhledem k odvozovacím pravidlům **(MP)**, **(NEC)** (a tedy i **(MON)** a **(CON)**).

V systému **K** ovšem nejsou platné některé vlastnosti, které bychom očekávali u (silné) znalosti. Systém **K** umožňuje agentovi např. vědět nepravdu, což je v rozporu s naší intuicí, že nelze vědět něco, co není pravda. Z tvrzení „Alenka ví, že prší“ nemůžeme jednoduše odvodit „Prší“. Nemožnost tohoto odvození je důsledkem neplatnosti axiomu **(T)**. Jak docílíme toho, aby platil? Tím, že omezíme množinu modelů $\mathcal{M}$, které budeme brát v potaz. A jak omezíme množinu těchto modelů? Tím, že na relaci $\mathcal{R}$ nasadíme určitá omezení. Konkrétněji, budeme chtít, aby relace $\mathcal{R}$ byla reflexivní, tj. aby každý svět byl dosažitelný ze sebe sama, formálněji:

- $\mathcal{R}(s, s)$ pro všechna $s \in S$.

Bude-li tedy relace $\mathcal{R}$ reflexivní, formule $K\varphi \to \varphi$ neboli axiom **(T)** bude opět platný. Obecně pak platí, že každý Kripkeho model $\mathcal{M}$, kde relace $\mathcal{R}$ je reflexivní, splňuje $\mathcal{M} \models K\varphi \to \varphi$. Dostali jsme tedy nový model znalosti s odlišnými vlastnosti, než měl původní systém **K**. Tomuto novému systému, který jsme získali rozšířením **K** o axiom **(T)**, budeme říkat systém **T**. Systém **T** je korektní a úplný vzhledem k množině všech Kripkeho modelů, ve kterých je relace $\mathcal{R}$ reflexivní.

Ovšem ani **T** nesplňuje všechny požadavky, které bychom mohli po našem modelu znalosti chtít. Další rozšířenou intuicí (ovšem zdaleka ne neproblematickou) je to, že pokud něco víme, pak také víme, že to víme, resp. jsme si toho vědomi (viz pozitivní introspekce výše). Touto vlastností avšak náš dosavadní systém **T** zatím neoplývá. Přesněji, formule $K\varphi \to KK\varphi$ není v **T** platná. To znamená, že v rámci systému **T** můžeme

něco vědět, aniž bychom si toho byli vědomi, konkrétněji, v systému **T** mohou současně platit formule $K\varphi$ a $\neg KK\varphi$. Co je potřeba udělat proto, abychom tento stav změnili? Jak už čtenář nejspíše správně tuší, cesta k úspěchu povede opět přes modifikaci vlastností relace $\mathcal{R}$. Tentokrát po ní budeme požadovat, aby byla tranzitivní, tj. aby splňovala následující podmínku:

- $\big(\mathcal{R}(s,t) \wedge \mathcal{R}(t,u)\big) \Rightarrow \mathcal{R}(s,u)$ pro všechna $s,t,u \in S$.

Slovní vyjádření: pokud je svět t dosažitelný ze světa s a svět u je dosažitelný ze světa t, pak je svět u dosažitelný ze světa s. Pokud bude relace $\mathcal{R}$ splňovat tuto podmínku, formule $K\varphi \rightarrow KK\varphi$ bude platná. Obecně pak platí, že každý Kripkeho model $\mathcal{M}$, kde $\mathcal{R}$ je tranzitivní, splňuje $\mathcal{M} \models K\varphi \rightarrow KK\varphi$. Systém **T** rozšířený o formuli $K\varphi \rightarrow KK\varphi$ neboli axiom **(4)** se nazývá **S4**. Stejně jako **K** a **T** i **S4** je korektní a úplný. Agent systému **S4** je schopen znát jen pravdivé formule (axiom **(T)**) a rovněž si je vědom všech svých znalostí (axiom **(4)**).

Další vlastností, kterou bychom po našem modelu znalosti mohli vyžadovat, je již diskutovaná negativní introspekce. Posouzení adekvátnosti tohoto axiomu přenecháme čtenáři. Formálně se jedná o formuli $\neg K\varphi \rightarrow K\neg K\varphi$. Tato formule je ale v systému **S4** zatím neplatná, a bude jej tedy třeba opět upravit prostřednictvím změn podmínek na relaci $\mathcal{R}$. Po ní budeme tentokrát chtít, aby se chovala jako relace ekvivalence. Relace ekvivalence, jak už víme, je relace $\mathcal{R}$, která splňuje následující tři podmínky:

- $\mathcal{R}(s,s)$ pro všechna $s \in S$,
- $\mathcal{R}(s,t) \Leftrightarrow \mathcal{R}(t,s)$ pro všechna $s,t \in S$,

- $\big(\mathcal{R}(s,t) \wedge \mathcal{R}(t,u)\big) \Rightarrow \mathcal{R}(s,u)$ pro všechna $s,t,u \in S$.

relace $\mathcal{R}$ je už reflexivní a tranzitivní, tudíž zbývá jen přidat podmínku symetrie.[25]

Systém **S4** rozšířený o formuli $\neg K\varphi \rightarrow K\neg K\varphi$ neboli axiom **(5)** se nazývá **S5**. Systém **S5** má největší expresivní sílu, neboť v sobě zahrnuje vlastnosti všech doposud diskutovaných systému, tj. **K**, **T** a **S4**. Stejně jako systémy **K**, **T** a **S4** i **S5** je korektní a úplný, a to ve všech modelech $\mathcal{M}$, kde $\mathcal{R}$ je relace ekvivalence. Tím jsme se dostali zpět k našemu výchozímu modelu znalosti SEL, který byl vystavěn právě na **S5** axiomech.

V současnosti nejrozšířenějšími modely znalosti v rámci SEL jsou **S4** a **S5**, přičemž **S4** odpovídá modelu znalosti, který Hintikka předkládá v *K&B*, a **S5** je rozšířený především mezi informatiky, právě kvůli svým introspektivním vlastnostem. Frans Voorbraak mluví o systému **S5** jako o systému modelujícím *objektivní znalost*, tj. takovou znalost, která může být připsána jak racionálním (lidským) agentům, tak i teploměrům, počítačovým systémům, zkrátka čemukoli, u čeho lze předpokládat jakýsi interní informační stav. Jinak řečeno, **S5**-znalost je běžně užívána jako externě připisovaná znalost, tj. taková znalost, která u agenta nevyžaduje uvažovací dovednosti.

4.3.2 Doxastická logika

Nyní se krátce zmíníme o doxastické logice neboli o logice přesvědčení. Přestože se označení doxastická logika užívá velmi zřídka a jen málo

[25] Ale pozor, platnost axiomu **(5)** nekoresponduje s tím, že $\mathcal{R}$ je symetrická, ale s tím, že je eukleidovská.

autorů důsledně rozlišuje mezi ní a epistemickou logikou, bude užitečné si ji představit zvlášť.

Čím se liší znalost od přesvědčení? Zatímco znalost musí být vždy pravdivá, v případě přesvědčení (věření, domnívání se…) se nic takového obecně nevyžaduje. Stručně řečeno, věřit můžeme, čemu chceme, ale pouze pravdu můžeme vědět. Tuto základní intuici využijeme k tomu, abychom vymezili hranici mezi doxastickou a epistemickou logikou v užším smyslu. Naším předpokladem bude tedy to, že jediným rozdílem mezi znalostí a přesvědčením je právě jejich závaznost vůči pravdivosti. Vzpomeňme si, že intuici, že znalost musí být pravdivá, jsme do systému SEL vnesli pomocí jistého axiomu, resp. podmínky na relaci $\mathcal{R}$. Máme tedy velmi ulehčenou situaci, neboť nemusíme logiku přesvědčení budovat celou od začátku, ale stačí jen identifikovat daný axiom a odstranit jej.

Hledaným axiomem je **(T)**, budeme tedy chtít popřít platnost formule $K\varphi \to \varphi$. Jak toho dosáhneme? Bude opět potřeba modifikovat vlastnosti relace $\mathcal{R}$. Vzpomeneme-li si, platnost této formule zaručovala reflexivita relace $\mathcal{R}$:

- $\mathcal{R}(s, s)$ pro všechna $s \in S$.

Pokud se této podmínky zbavíme, formule **(T)** přestává platit a my získáváme požadovanou doxastickou logiku. Odmítnutím axiomu **(T)** vlastně umožníme agentovi to, aby nemusel do svých úvah zahrnovat aktuální svět. Samozřejmě mu ale nic nebrání v tom, kdyby tak učinit chtěl.

Obecně řečeno, logiky přesvědčení jsou logiky znalosti bez axiomu **(T)**. Abychom dali lépe najevo, že nyní pracujeme s doxastickou logikou, obecný operátor K nahradíme operátorem B. Změna operátoru ovšem

není nezbytná a slouží spíše jen pro přehlednost, neboť jak v případě doxastické, tak i epistemické logiky v užším smyslu jsou to axiomy, které vymezují interpretaci modálního operátoru. Formuli $B\varphi$ budeme číst jako „agent je přesvědčen, že φ“, „agent věří, že φ“, „agent se domnívá, že φ“ atd. Pokud převedeme systém **S5** na doxastickou logiku, tj. vynecháme axiom **(T)**, získáme systém **K45**, který je definován následovně:

Axiomy

(P) Všechny tautologie výrokové logiky

(K^B) $B\varphi \wedge B(\varphi \to \psi) \to B\psi$

(4^B) $B\varphi \to BB\varphi$

(5^B) $\neg B\varphi \to B\neg B\varphi$

Odvozovací pravidla

(MP) $\dfrac{\varphi \quad \varphi \to \psi}{\psi}$

(NEC^B) $\dfrac{\varphi}{B\varphi}$

Systém **K45** je korektní a úplný vzhledem k množině všech Kripkeho modelů, kde relace dosažitelnosti $\mathcal{R}$ je tranzitivní a eukleidovská, tj. relace $\mathcal{R}$ splňuje následující dvě podmínky:

- $\big(\mathcal{R}(s,t) \wedge \mathcal{R}(t,u)\big) \Rightarrow \mathcal{R}(s,u)$ pro všechna $s,t,u \in S$,
- $\big(\mathcal{R}(s,t) \wedge \mathcal{R}(s,u)\big) \Rightarrow \mathcal{R}(t,u)$ pro všechna $s,t,u \in S$.

S tranzitivitou jsme si již setkali dříve, a tak jen slovní vyjádření *eukleidovské* podmínky: pokud jsou světy t a u dosažitelné ze světa s, pak je svět u také dosažitelný ze světa t.

Přestože se po přesvědčeních nevyžaduje pravdivost, obecně bychom určitě chtěli, aby byly alespoň konzistentní. Racionální agent by zkrátka neměl zastávat protichůdná, navzájem si odporující přesvědčení či přímo věřit kontradikci. Pokud je Alenka např. přesvědčena o tom, že prší, pak nemůže současně věřit tomu, že neprší.[26] Chceme tedy, aby platila formule:

(D) $B\varphi \rightarrow \neg B\neg\varphi$

Axiom **(D)** neboli D-axiom,[27] někdy také nazývaný jako *axiom konzistence*, zaručuje bezrozpornost přesvědčení agenta, tj. zabraňuje tomu, aby mohl věřit jak formuli φ, tak i zároveň její negaci. Krátce řečeno, axiom **(D)** dohlíží na to, aby si agent neprotiřečil. Proč jsme tento axiom nezavedli už v logice znalosti? Copak znalosti nemusí být konzistentní? Samozřejmě musí, nicméně jejich konzistenci garantuje už axiom **(T)**, proto není důvod zavádět axiom **(D)** do epistemické logiky v užším smyslu.

Pokud systém **K45** rozšíříme o axiom **(D)**, dostaneme systém **KD45**, někdy nazývaný také jako slabá **S5**. Proč slabá **S5**? Protože **KD45** se někdy považuje za logiku tzv. racionálního či silného (ospravedlněného) přesvědčení, která ovšem není pravou znalostí (nemusí být pravdivé).[28]

[26] Na to, že situace ovšem zdaleka nemusí být vždy tak jednoduchá a že racionalita nemusí jít vždy ruku v ruce s konzistentností, upozornil např. David Makinson tzv. paradoxem předmluvy, viz [Makinson, 1965]. Za tuto poznámku děkuji Igoru Sedlárovi. K tématu nekonzistentních přesvědčení se ještě vrátíme v sekci 5.2.1.

[27] D-axiom proto, že tímto axiomem se vyznačuje deontická logika, tj. logika zabývající se takovými pojmy jako nařízení, povolení atp.

[28] Byl to pravděpodobně Wolfgang Lenzen [Lenzen, 1978], kdo první zavedl rozlišení

KD45 je korektní a úplný vzhledem k množině všech Kripkeho modelů, kde je relace dosažitelnosti $\mathcal{R}$ tranzitivní, eukleidovská a seriální. Relace $\mathcal{R}$ tedy splňuje následující tři podmínky:

- $\big(\mathcal{R}(s,t) \wedge \mathcal{R}(t,u)\big) \Rightarrow \mathcal{R}(s,u)$ pro všechna $s,t,u \in S$,
- $\big(\mathcal{R}(s,t) \wedge \mathcal{R}(s,u)\big) \Rightarrow \mathcal{R}(t,u)$ pro všechna $s,t,u \in S$,
- pro všechna $s \in S$ existuje $t \in S$ takové, že $\mathcal{R}(s,t)$.

Seriálnost zachycuje tu skutečnost, že pokud jsou agentova přesvědčení zachycena nějakým světem, pak musí být konzistentní.

Shrnutí. Ukázali jsme si, že přidáváním dodatečných axiomů spolu s odpovídajícími úpravami na relaci $\mathcal{R}$ se můžeme od základního systému **K** dopracovat k silnějším logikám postihujícím další vlastnosti, které se tradičně znalosti připisují. Čím více axiomů přidáme, tím lepší bude mít agent přístup ke svému epistemickému univerzu (např. **(4)** mu umožní znát vlastní znalosti atd.). Obvykle bývá systém **K** rozšiřován o následující axiomy (kvůli lepší přehlednosti nyní použijeme jednotně symbol K):

(T) $K\varphi \to \varphi$

(D) $K\varphi \to \neg K \neg \varphi$

(4) $K\varphi \to KK\varphi$

mezi tzv. *silným* a *slabým* přesvědčením. Významový rozdíl mezi slabým a silným přesvědčení odpovídá víceméně rozdílu mezi českým „domnívat se“, „věřit“ na jedné straně a „být přesvědčen“ na straně druhé. Čtenář jistě cítí rozdíl mezi „Domnívám se, že to mám správně“ a „Jsem přesvědčen o tom, že to mám správně“. Rozdíl tedy spočívá v míře důvěry, kterou agent vkládá do svých přesvědčení.

(5) $\neg K\varphi \to K\neg K\varphi$

Nejčastěji se tak můžeme setkat s následujícími systémy epistemické logiky:

- **T** je systém **K** rozšířený o axiom **(T)**,
- **S4** je systém **T** rozšířený o axiom **(4)**,
- **S5** je systém **S4** rozšířený o axiom **(5)**,
- **KD45** je systém **S5** rozšířený o axiom **(D)**, bez axiomu **(T)**.

Pod SEL budeme řadit všechny tyto klasické modely znalosti a přesvědčení, konkrétně **K**, **T**, **S4**, **S5** a **KD45**, ovšem, neuvedeme-li výslovně jinak, budeme v pozadí SEL předpokládat systém **S5**, přičemž za kanonickou doxastickou logiku budeme považovat **KD45**.

Jak jsme si ukázali, platnost jednotlivých axiomů se odvíjí od toho, jaké podmínky budeme u relace $\mathcal{R}$ předpokládat, resp. od omezení množiny modelů $\mathcal{M}$. Postupně jsme se seznámili s těmito vlastnostmi relace $\mathcal{R}$:

- $\mathcal{R}$ je *reflexivní*, jestliže $\mathcal{R}(s,s)$ pro všechna $s \in S$,
- $\mathcal{R}$ je *tranzitivní*, jestliže $\big(\mathcal{R}(s,t) \wedge \mathcal{R}(t,u)\big) \Rightarrow \mathcal{R}(s,u)$ pro všechna $s,t,u \in S$,
- $\mathcal{R}$ je *symetrická*, jestliže $\mathcal{R}(s,t) \Leftrightarrow \mathcal{R}(t,s)$ pro všechna $s,t \in S$,
- $\mathcal{R}$ je *eukleidovská*, jestliže $\big(\mathcal{R}(s,t) \wedge \mathcal{R}(s,u)\big) \Rightarrow \mathcal{R}(t,u)$ pro všechna $s,t,u \in S$,
- $\mathcal{R}$ je *seriální*, jestliže pro všechna $s \in S$ existuje $t \in S$ takové, že $\mathcal{R}(s,t)$,
- $\mathcal{R}$ je *relace ekvivalence*, jestliže $\mathcal{R}$ je reflexivní, symetrická a tranzitivní.

Určitý model $\mathcal{M}$ pak můžeme nazvat reflexivní (symetrický,...) tehdy, jestliže relace $\mathcal{R}$ v $\mathcal{M}$ je reflexivní (symetrická,...).

Vztahy mezi jednotlivými axiomy a vlastnostmi relace $\mathcal{R}$ můžeme shrnout v následující tabulce:

Axiom	**Formule**	**Restrikce na relaci** $\mathcal{R}$
(K)	$K\varphi \wedge K(\varphi \to \psi) \to K\psi$	–
(T)	$K\varphi \to \varphi$	reflexivní
(D)	$K\varphi \to \neg K\neg\varphi$	seriální
(4)	$K\varphi \to KK\varphi$	tranzitivní
(5)	$\neg K\varphi \to K\neg K\varphi$	eukleidovská

Všechny systémy definované těmito axiomy (tj. **K** až **S5** a **KD45**) jsou korektní a úplné, tj. množina teorémů odpovídá množině platných formulí (tautologií) a naopak.

4.3.3 Mezi S4 a S5

Doposud jsme se seznámili jen s těmi základními modely znalosti, existuje ale samozřejmě i celá řada dalších, ne tak rozšířených systémů. Jak už bylo naznačeno, systém **S5** je mnohými považován jako příliš silný (právě kvůli problematickému axiomu negativní introspekce), kdežto **S4** zase jako příliš slabý. Rozumným východiskem se tedy zdá nalezení takového systému, který leží někde mezi **S4** a **S5**. Mezi takové logiky patří např. **S4.2**, **S4.3** a **S4.4**. Nebudeme se jimi zabývat moc do hloubky, pouze se ve stručnosti seznámíme s jejich charakteristickými axiomy. Pokud k systému **S4** přidáme axiom **(0.2)**:

(0.2) $\neg K\neg K\varphi \to K\neg K\neg\varphi$

dostaneme systém **S4.2**. Axiom **(0.2)** je někdy označován jako *axiom konvergence* nebo G-axiom. Dle Franse Voorbraaka modeluje systém

S4.2 zdůvodněné pravdivé přesvědčení (angl. justified true belief neboli JTB)[29] a Wolfgang Lenzen se domnívá, že jde o pravou logiku znalosti.[30]

Přidáme-li k systému **S4** axiom **(0.3)**:

(0.3) $K(K\varphi \to KK\psi) \vee K(K\psi \to K\varphi)$

získáme systém **S4.3**, který navrhl k posilnění **S4** van der Hoek,[31] a systém **S4.4** se zavádí pomocí axiomu **(0.4)**:

(0.4) $\varphi \to (\neg K\neg K\varphi \to K\varphi)$

Dle Lenzena definuje systém **S4.4** logiku pravdivého přesvědčení.

Situaci mezi systémy **S4** a **S5** můžeme shrnout v následující tabulce:

Systém	Formule	Axiomy
S4		
S4.2	$\neg K\neg K\varphi \to K\neg K\neg\varphi$	**(K)**, **(T)**, **(4)**, **(0.2)**
S4.3	$K(K\varphi \to KK\psi) \vee K(K\psi \to K\varphi)$	**(K)**, **(T)**, **(4)**, **(0.3)**
S4.4	$\varphi \to (\neg K\neg K\varphi \to K\varphi)$	**(K)**, **(T)**, **(4)**, **(0.4)**
S5		

Tím máme definování SEL za sebou. V následující sekci ji otestujeme v praxi, tj. při rozboru konkrétního případu.

4.3.4 Aplikace SEL

Pokusme se v rámci SEL analyzovat situaci, kdy Alenka uvažovala o počasí. Připomeňme si v krátkosti, o co v daném příkladu šlo. Alenka se

[29] Srov. [Voorbraak, 1992].

[30] Srov. [Lenzen, 1979].

[31] Srov. [van der Hoek, 1993].

procházela po Brně a přemýšlela nad tím, jestli zrovna prší nebo neprší v Praze a Ostravě. Z toho jsme usoudili, že Alenka uvažuje dohromady o čtyřech možných světech:

- *svět*$_1$ = {prší v Praze, prší v Ostravě},
- *svět*$_2$ = {prší v Praze, neprší v Ostravě},
- *svět*$_3$ = {neprší v Praze, prší v Ostravě},
- *svět*$_4$ = {neprší v Praze, neprší v Ostravě}.

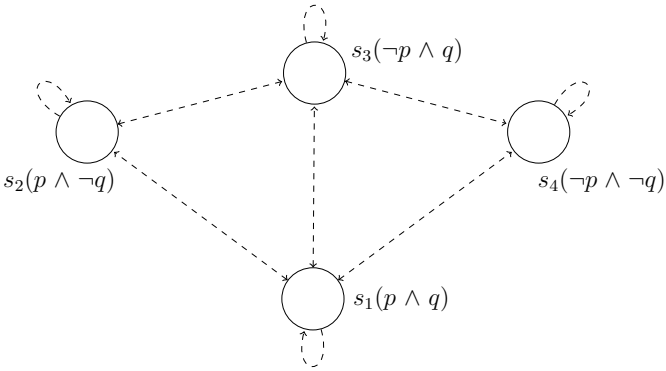
Tím jsme vyčerpali všechny myslitelné varianty, které mohou nastat. Jinými slovy, Alenčina množina epistemických alternativ je v tomto případě tvořeno čtyřmi možnými světy s_1, s_2, s_3 a s_4.

Nyní přistupme k formálnímu zachycení této situace. Nechť jazyk $\mathcal{L}$ obsahuje pouze dvě atomické formule $\mathcal{P} = \{p, q\}$, které po řadě označují tvrzení „v Praze prší“ a „v Ostravě prší“, a $\mathcal{M}$ je Kripkeho model tvaru $\mathcal{M} = \langle S, \pi, \mathcal{R}\rangle$, přičemž $S = \{s_1, s_2, s_3, s_4\}$ a p je pravdivé ve světech $\{s_1, s_2\}$, nepravdivé ve světech $\{s_3, s_4\}$ a q je pravdivé ve světech $\{s_1, s_3\}$, nepravdivé ve světech $\{s_2, s_4\}$. Formálněji:

- pro s_1 platí, že $\pi(s_1)(p) = \texttt{true}$ a $\pi(s_1)(q) = \texttt{true}$,
- pro s_2 platí, že $\pi(s_2)(p) = \texttt{true}$ a $\pi(s_2)(q) = \texttt{false}$,
- pro s_3 platí, že $\pi(s_3)(p) = \texttt{false}$ a $\pi(s_3)(q) = \texttt{true}$,
- pro s_4 platí, že $\pi(s_4)(p) = \texttt{false}$ a $\pi(s_4)(q) = \texttt{false}$.

Alenčina množina epistemických alternativ tedy vypadá takto:

$$\{\{p, q\}, \{p, \neg q\}, \{\neg p, q\}, \{\neg p, \neg q\}\}.$$



Obrázek 4.8: Alenčiny epistemické alternativy

Tuto situaci můžeme znázornit v grafu (viz obr. 4.8).

Graf 4.8 znázorňuje čtyři možné alternativy, jak by se to mohlo mít s deštěm v Praze a Ostravě. Agent neví, jaké je tam počasí, proto považuje všechny čtyři světy za možné. Jinými slovy, agent s informacemi, které má k dispozici, od sebe nedokáže jednotlivé světy rozlišit. Agent si je samozřejmě moc dobře vědom toho, že světy s_1 až s_4 jsou odlišné, koneckonců, v některých prší, v jiných ne, ale nedokáže rozeznat, který z nich je ten aktuální.[32] Jinak řečeno, nemá dostatek údajů k tomu, aby mohl určit, který ze světů $\{s_1, s_2, s_3, s_4\}$ odpovídá skutečnosti.[33] To znamená,

[32] Přestože způsob, jakým jsme graf sestrojili, naznačuje, že aktuálním světem je svět s_1, nesmíme zapomínat na to, že tato skutečnost zůstává agentovi utajena. Rovněž stojí za poznámku, že v literatuře je zvykem používat v případě systému **S5** zjednodušené zobrazení, při kterém se vynechávají šipky reprezentující reflexivitu a zbylé šipky jsou nahrazeny neorientovanými hranami.

[33] Proč jsme do grafu nezanesli také Brno, resp. stav brněnského počasí? To proto, že na grafu jsme se rozhodli znázornit pouze ty formule, kterými se od sebe jednotlivé světy liší. Pokud bychom chtěli být skutečně důslední, mohli bychom samozřejmě přidat i formuli $\neg r$, která by označovala fakt „nepršíˇ v Brně“ (= „v Brně je slunný den“). Tu bychom ovšem museli připsat ke každému ze světů s_1 až s_4, neboť, jak už

že nedokáže rozlišit aktuální svět od ostatních možných světů a jsou pro něj tedy epistemicky nerozlišitelné. Analogicky by pak proběhla analýza i ostatních příkladů.

4.3.5 Systémy kombinující znalost a přesvědčení

Systémy pro znalost a přesvědčení jsme zatím brali vždy jen odděleně, tudíž tvrzení jako např.:

(j) Alenka ví, že prší, a je přesvědčena, že je pátek.

unikala naší analýze. Není ovšem žádný problém zavést takový systém, který obsahuje oba operátory, tj. K i B, s odpovídajícími axiomy a odvozovacími pravidly, což nám následně umožňuje tvrzení výše zachytit jako:

(j') $Kp \wedge Bq$

Takový systém by nám sice již umožnil analyzovat výše představené tvrzení (j), nicméně významy obou epistemických operátorů zůstávají v podstatě izolovány ve svých domovských systémech, tj. po řadě **S5** a **KD45**. Obecná intuice je avšak taková, že znalost a přesvědčení spolu určitým způsobem souvisí. Uvážíme-li např. klasickou JTB definici znalosti, přesvědčení se stává dokonce nutným předpokladem znalosti. Náš systém bude tedy vyžadovat určité propojovací axiomy, které nám umožní tento vztah zachytit. S takovými axiomy přišli Sarit Kraus a Daniel Lehmann.[34] Uvažme následující formuli:

víme, znalost je definována jako pravda ve všech světech, které agent považuje za možné. Byl by to tedy jen údaj navíc, který by nijak nezvyšoval celkovou informativnost grafu, a proto jej vynecháváme.

[34] Srov. [Kraus & Lehmann, 1988].

(KB) $K\varphi \rightarrow B\varphi$

Slovně: pokud agent ví, že φ, pak agent také věří, že φ. Axiom **(KB)** se opírá o výše zmíněné tradiční pojetí znalosti jako pravdivého zdůvodněného přesvědčení. Uvažme další axiom:

(BK) $B\varphi \rightarrow KB\varphi$

Slovní vyjádření: pokud agent věří, že φ, pak také ví, že věří, že φ. Tento axiom **(BK)** zachycuje explicitnost našich přesvědčení, tj. tu skutečnost, že nemáme žádná neuvědomělá přesvědčení. Tento axiom se už nemusí jevit tak neproblematický jako **(KB)**. Příkladem by mohly být např. nějaké předsudky, kterých si nejsme ani vědomi. Obecně ale budeme předpokládat oprávněnost tohoto principu. Třetím a posledním axiomem bude:

(BB) $B\varphi \rightarrow BK\varphi$

Slovně: pokud agent věří, že φ, pak také věří, že ví, že φ. Úvaha v pozadí je zhruba zachycena následujícím monologem agenta: „Přece bych nevěřil něčemu, co není pravda, takže pokud něčemu věřím, pak to musí být pravda“. Chybnost této úvahy je zřejmá. Axiom **(BB)**, někdy nazývaný jako *Moorův princip*, bychom mohli popsat jako axiom epistemické tvrdohlavosti či ješitnosti racionálního agenta vycházející z té skutečnosti, že nikdo se nechce záměrně mýlit. Jinými slovy, nikdo úmyslně nepovažuje nepravdivá přesvědčení za pravdivá, tj. pokud agent něčemu věří, pak také předpokládá, že je to pravda. Axiom **(BB)** nelze uznat za obecně platný princip, ne vždy totiž máme tolik důvěry ve svá přesvědčení a nic není kontradiktorické na tvrzeních jako „Věřím, že jsem

zamknul dveře, ale jistě to nevím“ či „Jsem přesvědčený o tom, že mám správný výsledek, ale můžu se mýlit“ apod.

Kraus a Lehmann si dále všimli, že přijetí axiomu **(BB)** do jejich systému (viz níže) má mnohem závažnější následky, než by se na první pohled mohlo zdát. Konkrétněji, tento axiom by umožnil odvození formule $K\varphi \leftrightarrow B\varphi$, čímž by se zhroutila hranice mezi znalostí a přesvědčením, což je nepochybně nežádoucí důsledek.[35]

Nyní přistupme k formálnímu zachycení tohoto systému, přičemž se omezíme jen na ty části, ve kterých se liší od SEL. Jak si lze snadno domyslet, půjde o systém, který bude obsahovat dva operátory zastupující dvě odlišné modality (tj. znalost a přesvědčení), a tedy i dvě odpovídající relace $\mathcal{R}$.

Nechť SEL^{KB} je logika, která obsahuje axiomy systémů **S5** a **KD45** doplněné o následující propojovací axiomy, které zprostředkovávají interakci mezi logikou znalosti a logikou přesvědčení.

(KB) $K\varphi \to B\varphi$

(BK) $B\varphi \to KB\varphi$

Definování sémantiky omezíme jen na popis odpovídajícího Kripkeho modelu.

Nechť $\mathcal{L}$ je jazyk výrokové logiky a $\mathcal{P}$ množina atomických formulí, pak Krauss-Lehmannův model $\mathcal{M}$ má tvar $\langle S, \pi, \mathcal{R}^K, \mathcal{R}^B \rangle$, kde:

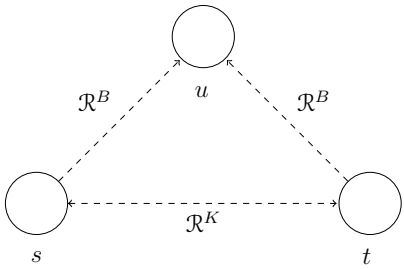
- S je neprázdná množina možných světů,
- π je funkce, která přiřazuje pravdivostní hodnotu atomickým formulím v možných světech S, tj. $\pi : S \mapsto (\mathcal{P} \mapsto \{\texttt{true}, \texttt{false}\})$,

[35] Návrhy možných řešení se zde zabývat nebudeme, ale je možné je naleznout např. ve [van der Hoek, 1991] či [Voorbraak, 1991].

- $\mathcal{R}^K$ je relace dosažitelnosti vzhledem ke znalosti,
- $\mathcal{R}^B$ je relace dosažitelnosti vzhledem k přesvědčení,

přičemž $\mathcal{R}^K$ je relací ekvivalence a $\mathcal{R}^B$ je seriální relace taková, že $\mathcal{R}^B \subseteq \mathcal{R}^K$ a dále platí, že $\mathcal{R}^K \in (s,t)$ a $\mathcal{R}^B \in (t,u)$ implikují $\mathcal{R}^B \in (s,u)$ pro všechna $s,t,u \in S$.

Jak můžeme vidět, znalost je zde tedy pojímána jako **S5**-znalost (v důsledku toho, že relace $\mathcal{R}^K$ je relací ekvivalence) a přesvědčení jako **KD45**-přesvědčení. Seriálnost, jak už víme, zachycuje tu skutečnost, že agent vždy považuje alespoň jednu epistemickou alternativu za možnou. Podmínka $\mathcal{R}^B \subseteq \mathcal{R}^K$ zachycuje tu skutečnost, že pokud agent něco ví, pak tomu také věří. Poslední podmínka říká, že pokud jsou světy s a t možné na základě znalosti agenta a jestliže agent věří ve světě t, že u je možné, pak také věří ve světě s, že u je možné, což je právě důsledkem toho, že od sebe světy s a t nedokáže rozlišit. To tedy znamená, že každý svět, který je propojený relací $\mathcal{R}^K$, sdílí stejné možné světy vzhledem k přesvědčení (viz obr. 4.9).



Obrázek 4.9: Znalost a přesvědčení

4.3.6 Systémy s více agenty

Kvůli jednoduchosti jsme doposud pracovali se systémem s jediným agentem, v našem případě s Alenkou. Není třeba dodávat, že tento přístup je značně limitující a ani zdaleka neodpovídá naší běžné epistemické praxi, kde vzájemný kontakt dvou a více agentů je doslova každodenní záležitostí. V takovýchto situacích je třeba brát v potaz nejen to, co agent ví o světě, ale i to, co ví o znalostech jiných agentů, tj. o tom co ví a neví ostatní.

Tento typ úvah je v přirozeném jazyce velmi nepřehledný a tvrzení jako „Alenka neví, jestli Bedřich ví, že Alenka ví, že Bedřich ví, že Cyril snědl její čokoládové sušenky“ jsou toho skvělým příkladem. Není třeba dodávat, že tento typ uvažování hraje důležitou roli nejen v rámci epistemické logiky, ale také v teorii her (vězňovo dilema), ekonomii a obecně všude, kde dochází k interakci dvou a více racionálních agentů. Monoagentní systémy nás tedy velmi omezují v počtu možných situací, které můžeme modelovat. Uvažme např. následující tvrzení:

(d) Alenka ví, že Bedřich ví, že prší.

V rámci SEL, tj. v systému s jediným agentem, bychom mohli (d) zachytit pouze jako:

(d’) Alenka ví, že q

formálněji jako:

(d”) $K_A q$

Tím ale přijdeme o mnoho zajímavých informací. Alenčina znalost, že „Bedřich ví, že prší“ není totiž jen tak ledajakou znalostí. Je to znalost

o znalosti nějakého jiného agenta, jmenovitě Bedřicha. Chceme tedy takový systém, který by byl schopen tuto informaci zachytit. Naším cílem je tak dospět k následujícímu schématu:

(f') A ví, že B ví, že p

formálněji pak:

(f") $K_A K_B p$

Už na první pohled je zřejmé, že analýza (f") s sebou nese mnohem více informací než (d"). Zatímco z formule (d") může vyčíst pouze to, že Alenka ví, že platí nějaké q, formule (f") prozrazuje i to, že Alenčina znalost se týká znalosti někoho jiného. Tato informace nám nebyla dříve přístupná. A co je mnohem zajímavější, formule (f") nám umožňuje odvodit i dodatečnou formuli, a to

(g') $K_A p$

Jinými slovy, z toho, že „Alenka ví, že Bedřich ví, že prší", můžeme odvodit i to, že „Alenka ví, že prší". Takovéto odvození není v rámci SEL možné. Naštěstí ale není problém náš formální model znalosti SEL upravit tak, aby mohl agentů zahrnout více.

Všimněme si, že v situacích jako (d) vlastně zachycujeme epistemické postoje agenta, které se vztahují k epistemickým postojům jiných agentů. Popisujeme tak jakousi intersubjektivní metaznalost, tj. znalost o znalostech jiných agentů. Modelování těchto znalostí vyššího řádu je právě úkolem *multiagentních systémů*, tj. systémů, které pracují s více než

jedním agentem.[36] Příkladem mohou být např. systémy pro správu počítačových sítí, software na řízení letového provozu, ale také skupina dětí hrající si na pískovišti atp. Nyní tedy rozšíříme SEL na multiagentní systém, který nám umožní analyzovat (d).

4.3.6.1 Rozšíření SEL na SEL^n

Ačkoliv byla SEL původně navržena jen pro popis jediného agenta, drobnými úpravami lze tento nedostatek snadno napravit a vytvořit tak ze SEL systém, který je schopný zachytit libovolný počet agentů. Takový systém budeme označovat jako SEL^n. Prvním krokem je to, že rozšíříme jazyk SEL o n operátorů znalosti K, přičemž n je počet agentů. Na každého agenta tak připadne jeden znalostní operátor. Analogicky pak do sémantiky zavedeme n relací epistemické dosažitelnosti $\mathcal{R}$ popisující právě chovaní n operátorů K. Výsledný multiagentní systém SEL^n bude tedy pro n agentů obsahovat n oddělených znalostních operátorů. Jinak řečeno, SEL pro n agentů je složena z n kopií dané SEL, přičemž v rámci zjednodušení se předpokládá, že agenti jsou homogenní, tj. že mohou být popsáni stejnou logikou. SEL^n je tedy *multimodální logikou*.

Nechť $\mathcal{L}^n$ je jazyk výrokové modální logiky rozšířený o n operátorů K, $\mathcal{P}$ neprázdná množina atomických formulí a $Agt = \{1, \ldots, n\}$ množina agentů, pak Kripkeho model s n agenty je model $\mathcal{M}$ tvaru $\langle S, \mathcal{R}_1, \ldots, \mathcal{R}_n, \pi \rangle$, kde:

- S je neprázdná množina možných světů,
- $\mathcal{R}_i \subseteq S \times S$ pro $1 \leq i \leq n$, je relace dosažitelnosti pro agenty z Agt,

[36] Multiagentní systém je termín, který má v informatice svůj přesný význam, nicméně v rámci epistemické logiky jím budeme obecně označovat každý systém se dvěma a více agenty.

- π je funkce, která přiřazuje pravdivostní hodnotu atomickým formulím v možných světech S, tj. $\pi : S \mapsto (\mathcal{P} \mapsto \{\mathsf{true}, \mathsf{false}\})$.

Sémantika je definována obdobným způsobem jako v SEL a platné formule modelu jsou určeny axiomy **S5** (tj. $\mathcal{R}_i$ je relace ekvivalence). Jediným rozdílem je to, že pro každého agenta vezmeme právě jednu **S5** axiomatizaci. Např. formuli $K_i \neg K_j p$ pak budeme číst jako „agent i ví, že agent j neví, že p".

V SEL^n už můžeme zachytit tvrzení (d):

(d) Alenka ví, že Bedřich ví, že prší.

jako požadované:

(f") $K_i K_j p$

SEL^n, v podobě v jaké jsme si jí zatím představili, ovšem není moc zajímavou logikou. Jsme v ní sice schopni vyjádřit znalost agenta o znalostech jiných agentů, ale nedokážeme např. vyjádřit to, že všichni agenti určité skupiny sdílí nějakou znalost nebo že všichni vědí to stejné. Abychom mohli zachytit i tyto situace, ve kterých vlastně dochází k interakci znalostí jednotlivých agentů, je třeba naši SEL^n rozšířit dále, a to zavedením dvou nových pojmů *společné* (angl. common) a *distribuované* (angl. distributed) znalosti,[37] které nám pomohou dále rozšířit potenciál SEL^n.

[37] Srov. [Halpern & Moses, 1985], [Meyer, 2001].

4.3.6.2 Společná znalost

Společná znalost je taková znalost, kterou vzájemně sdílí všichni členové skupiny, přičemž řekneme, že skupina má nějakou společnou znalost φ právě tehdy, když všichni ví, že φ, a všichni ví, že všichni ví, že φ atd.[38]

Nejdříve tedy budeme muset zachytit, co to znamená „všichni ví, že φ", což zapíšeme jako $E\varphi$ a jazyk SEL^n rozšíříme o příslušný symbol E. Je jasné, že celá skupina zná nějakou formuli φ právě tehdy, když φ zná každý člen dané skupiny. Tento nový typ znalosti tedy můžeme axiomatizovat pomocí následující formule:

(E) $E\varphi \leftrightarrow (K_1\varphi \wedge ... \wedge K_n\varphi)$

pro n agentů. Zbývá ještě definovat sémantiku nového operátoru:

- $(\mathcal{M}, s) \models E\varphi \Leftrightarrow (\mathcal{M}, t) \models \varphi$ pro všechna t taková, že $(s, t) \in \mathcal{R}_E$.

Relaci $\mathcal{R}_E$ budeme pojímat jako $\mathcal{R}_E = (\mathcal{R}_1 \cup \ldots \cup \mathcal{R}_n)$, tj. jako sjednocení všech relací $\mathcal{R}_i$ jednotlivých agentů. Epistemické alternativy všech agentů dané skupiny se tak spojí do jedné množiny, která bude představovat epistemické alternativy celé skupiny dohromady. Jde tedy o takovou znalost, kterou sdílí každý člen skupiny.

Operátor E dědí základní charakteristiky modálního operátoru systému **S5**, tj. **(K)**, **(NEC)** a **(T)** v odpovídajících variantách:

(K^E) $E\varphi \wedge E(\varphi \rightarrow \psi) \rightarrow E\psi$

(NEC^E) $\dfrac{\varphi}{E\varphi}$

[38] Jedním z prvních, kdo nabídl teorii společné znalosti, byl [Lewis, 1969], a to v rámci svého studia konvencí. Výklad zde vychází z [Meyer & Hoek, 1995].

(T^E) $E\varphi \to \varphi$

Přestože se pohybujeme v rámci axiomatiky **S5**, nejsou platné axiomy pozitivní a negativní introspekce. Důvodem je to, že pokud provedeme sjednocení relací $\mathcal{R}_i$, výsledná relace bude sice stále reflexivní, ale může přijít o tranzitivitu a eukleidovskost, a v důsledku tak i obě varianty introspekce. To ale odpovídá i našim intuicím ohledně znalosti ve skupině: není problém si např. představit takovou situaci, kdy všichni v kanceláři ví, že jejich kolega A má zápis v trestním rejstříku, a současně nemusí vědět, že to ví někdo další kromě nich. Jinak řečeno, každý může být přesvědčen o tom, že je to tajemství, které zná jen on či ona a nechává si ho pro sebe. Pokud tedy máme spolupracovníky B a C, přičemž p bude zastupovat „kolega A má zápis v trestním rejstříku“, můžeme to formálně vyjádřit jako $K_B p \wedge K_C p \wedge \neg K_B K_C p$, z čehož vyplývá právě $Ep \wedge \neg EEp$, tj. neplatnost introspekce.

Už máme definované formule tvaru $E\varphi$, takže se můžeme vrátit k definici společné znalosti. Jazyk SEL^n rozšíříme o další operátor C. Formule tvaru $C\varphi$ budeme číst jako „je společnou znalostí mezi agenty, že φ“. O nějaké formuli φ řekneme, že je společnou znalostí v určité skupině tehdy, když

1. všichni ví, že φ,
2. všichni ví, že všichni ví, že φ,
3. všichni ví, že všichni ví, že všichni ví, že ...

   atd. *ad infinitum*.[39]

[39] Proč je nutná induktivní definice společné znalosti lze naleznout např. v [Clark & Marshall, 1981].

Tato nekonečnost definice ovšem dosti stěžuje situaci, neboť náš jazyk $\mathcal{L}^n$ nepřipouští nekonečné formule, a tudíž ji nemůžeme formálně zachytit jednoduše jako nekonečnou konjunkci $E\varphi \wedge EE\varphi \wedge EEE\varphi \ldots$ atd. K tomu, abychom mohli interpretovat operátor společné znalosti C, budeme potřebovat zastřešující relaci $\mathcal{R}_C$, která bude spojovat jednotlivé agenty se společnou znalostí. Technicky je toho dosaženo tím, že budeme vyžadovat, aby relace $\mathcal{R}_C$ asociovaná s operátorem C byla tranzitivním uzávěrem sjednocení relací $\mathcal{R}_i$ jednotlivých agentů dané skupiny.

Nechť $\mathcal{L}^n$ je jazyk výrokové modální logiky, $\mathcal{P}$ neprázdná množina atomických formulí a $Agt = \{1, \ldots, n\}$ množina agentů, pak Kripkeho model s n agenty je model $\mathcal{M}$ tvaru $\langle S, \mathcal{R}_1, ..., \mathcal{R}_n, \mathcal{R}_C, \pi \rangle$, kde:

- S je neprázdná množina možných světů,
- $\mathcal{R}_i \subseteq S \times S$ pro $1 \leq i \leq n$, je relace dosažitelnosti (ekvivalence) pro agenty z Agt,
- $\mathcal{R}_C$ je tranzitivní uzávěr $\mathcal{R}_1 \cup ... \cup \mathcal{R}_n$ takový, že $(s,t) \in \mathcal{R}_C$ právě tehdy, když je tu posloupnost $s = s_0, s_1, ..., s_m = t$ taková, že $\mathcal{R}_E(s_i, s_{i+1})$ pro všechna $0 \leq i \leq m-1$,
- $\pi : S \mapsto (\mathcal{P} \mapsto \{\texttt{true}, \texttt{false}\})$.

Nyní definujeme sémantiku pro operátor C. Ta, jak už bylo řečeno, bude definována s pomocí zastřešující relace $\mathcal{R}_C$:

- $(\mathcal{M}, s) \models C\varphi \Leftrightarrow (\mathcal{M}, t) \models \varphi$ pro všechna t taková, že $(s,t) \in \mathcal{R}_C$.

Slovní vyjádření: formule φ je společnou znalostí ve skupině právě tehdy, když je φ pravdivá ve všech světech, které jsou dosažitelné relací $\mathcal{R}_C$,

přičemž relace $\mathcal{R}_C$ je tranzitivní uzávěr sjednocení všech relací $\mathcal{R}_i$ jednotlivých agentů dané skupiny. Vlastnosti relace $\mathcal{R}_C$ pro společnou znalost vyplývají z vlastností individuálních relací dosažitelnosti $\mathcal{R}_i$, ovšem nemusí to být nutně ty stejné. Jen pro připomenutí, relace $\mathcal{R}_i$ bereme jako relace ekvivalence. Systém pro společnou znalost je axiomatizován (společně s axiomy systému **S5**) následujícími formulemi:

(K^C) $C\varphi \wedge C(\varphi \to \psi) \to C\psi$

(T^C) $C\varphi \to \varphi$

(L) $C\varphi \to EC\varphi$

(E) $E\varphi \leftrightarrow (K_1\varphi \wedge ... \wedge K_n\varphi)$

(C$_{ind}$) $C(\varphi \to E\varphi) \to (\varphi \to C\varphi)$

význam **(K^C)** a **(T^C)** by měl být zřejmý, z axiomu **(L)** se dozvídáme, že pokud je nějaké φ společnou znalostí, pak všichni ví, že φ je společnou znalostí. Axiom **(E)** byl diskutovaný už výše a **(C$_{ind}$)** slouží k induktivnímu zachycení společné znalosti.

Zbývá už jen přidat C variantu necesitace:

(NEC^C) $\dfrac{\varphi}{C\varphi}$

Systém **S5**^{EC}, tj. multiagentní verze systému **S5** obsahující operátory E a C, je úplný a korektní.[40] Analogicky pak můžeme získat i **K**^{EC}, **T**^{EC} a **S4**^{EC}:

- **K**^{EC} je systém **K** rozšířený o axiomy **(K^C)**, **(T^C)**, **(K^E)**, **(E^C)**, **(C$_{ind}$)**, **(NEC^C)**,

[40] Srov. [Fagin et al., 1995].

- $\mathbf{T}^{EC}$ je systém **T** rozšířený o axiomy systému $\mathbf{K}^{EC}$,
- $\mathbf{S4}^{EC}$ je systém **S4** rozšířený o axiomy systému $\mathbf{T}^{EC}$,
- $\mathbf{S5}^{EC}$ je systém **S5** rozšířený o axiomy systému $\mathbf{S4}^{EC}$.

Multiagentní SEL rozšířenou o společnou znalost budeme označovat jako SEL^{nC}.

4.3.6.3 Distribuovaná znalost

Distribuovaná znalost φ, kterou budeme značit $D\varphi$, popisuje nějakou implicitní znalost φ ve skupině, která se může stát explicitní, pokud si členové skupiny navzájem sdělí to, co ví.

Jak zachytit koncept distribuované znalosti formálně? Vzpomeňme si, že znalost jednotlivého agenta v podstatě odpovídá množině jeho epistemických alternativ. Spojíme-li tedy tyto jejich epistemické alternativy, získáme požadovanou distribuovanou znalost. Že je tomu skutečně tak, se můžeme přesvědčit na jednoduchém příkladu. Předpokládejme, že Alenka ví, že:

(a_1) Flipper je delfín.

Její bratr Bedřich neví, že (a_1), ale naopak ví, že:

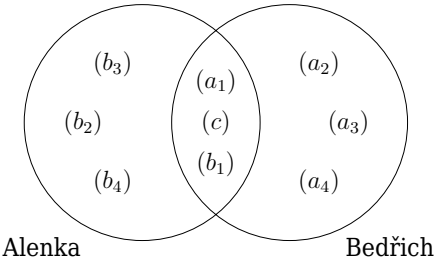
(b_1) Všichni delfíni jsou savci.

a (b_1) zase neví Alenka. V tomto okamžiku je tedy jejich distribuovanou znalostí:

(c) Flipper je savec.

Tvrzení (c) neví ani Alenka (neví, že delfíni jsou savci), ani Bedřich (neví, že Flipper je delfín), ale je to jejich distribuovaná neboli implicitní znalost, a dokud se o svoje informace sourozenci navzájem nepodělí, tak takovou i zůstane.

Technicky lze tento koncept znalosti definovat jako průnik znalostních množin (resp. epistemických alternativ) jednotlivých agentů (viz obr. 4.10 níže). Uvědomme si, že koncept distribuované znalosti nedělá nic jiného než to, že eliminuje epistemické alternativy každého agenta pomocí epistemických alternativ jiných agentů ve skupině. Alenka ví, že Flipper je delfín a nic dalšího. Množina jejich epistemických alternativ tedy bude obsahovat takové světy, kde budou např. pravdivá tvrzení jako (b_1) „Delfíni jsou savci“, (b_2) „Delfíni jsou paryby“, (b_3) „Delfíni jsou obojživelníci“, (b_4) „Delfíni jsou plazi“ atd. Dostane-li se k ní informace (b_1), může pak vyloučit ty alternativy, ve kterých jsou delfíni parybami, obojživelníky atd. Analogicky to pak probíhá v případě Bedřicha.



Obrázek 4.10: Distribuovaná znalost

Výsledkem takovéto komunikace bude tedy průnik množiny epistemických alternativ jednotlivých agentů, tj. $\mathcal{R}_D = (\mathcal{R}_1 \cap ... \cap \mathcal{R}_n)$, přičemž relaci $\mathcal{R}_D$ budeme říkat *relace distribuované znalosti*.

Nechť $\mathcal{L}^n$ je jazyk výrokové modální logiky, $\mathcal{P}$ neprázdná množina atomických formulí a $Agt = \{1, \ldots, n\}$ množina agentů, pak Kripkeho model s n agenty je model $\mathcal{M}$ tvaru $\langle S, \mathcal{R}_1, ..., \mathcal{R}_n, \mathcal{R}_D, \pi \rangle$, kde:

- S je neprázdná množina možných světů,
- $\mathcal{R}_i \subseteq S \times S$ pro $1 \leq i \leq n$, je relace dosažitelnosti pro agenty z Agt,
- $\mathcal{R}_D = (\mathcal{R}_1 \cap ... \cap \mathcal{R}_n)$,
- $\pi : S \mapsto (\mathcal{P} \mapsto \{\texttt{true}, \texttt{false}\})$.

Sémantiku operátoru D definujeme následovně:

- $(\mathcal{M}, s) \models D\varphi \Leftrightarrow (\mathcal{M}, t) \models \varphi$ pro všechna t taková, že $(s, t) \in \mathcal{R}_D$.

Čteme: φ je distribuovanou znalostí právě tehdy, když je formule φ platná ve všech světech t, které jsou dosažitelné relací distribuované znalosti $\mathcal{R}_D$. S operátorem D asociujeme tuto relaci $\mathcal{R}_D$. Jakými vlastnostmi se vyznačuje operátor D? Vedle axiomu **(K)**, **(T)**, obou introspekcí a necesitace v odpovídajících variantách, tj.:

(K^D) $D\varphi \wedge D(\varphi \rightarrow \psi) \rightarrow D\psi$

(T^D) $D\varphi \rightarrow \varphi$

(S4^D) $D\varphi \rightarrow DD\varphi$

(S5^D) $\neg D\varphi \rightarrow D\neg D\varphi$

(NEC^D) $\dfrac{\varphi}{D\varphi}$

splňuje také:

- $K_i\varphi \rightarrow D\varphi$

Systém obsahující takto definovaný operátor D je korektní a úplný. Stojí ještě za zmínku, že distribuovaná znalost ve skupině s jedním členem se redukuje na znalost tohoto agenta, tudíž platí také, že $D_i\varphi = K_i\varphi$ pro $n = 1$. Multiagentní systém SEL^n rozšířený o společnou i distribuovanou znalost budeme označovat jako SEL^{nCD}.

4.4 Rozšíření SEL na predikátovou logiku

Jak už bylo řečeno na začátku kapitoly, výklad SEL jsme omezili pouze na modální výrokovou logiku. To ovšem znamená, že v jazyce SEL nejsme schopni vyjádřit např. tvrzení, které je jednou z premis epistemického úsudku (2):

(2a) Alenka ví, že všichni lidé jsou smrtelní.

Už z dřívějška víme, že k adekvátní analýze (2a), resp. celého úsudku (2), budeme potřebovat predikátovou logiku. SEL vystavěnou na modální výrokové logice tedy rozšíříme o kvantifikátory, individuové konstanty a relace. Cílem následující sekce bude tedy definovat epistemickou logiku prvního řádu, kterou budeme označovat jako SEL*.

4.4.1 Výstavba SEL*

V této části si definujeme SEL*, tj. epistemickou logiku prvního řádu.

4.4.2 Syntax SEL*

Jazyk modální predikátové logiky s identitou se skládá z logických symbolů:

- spočetné nekonečné množiny individuových proměnných: $V = \{x, y, \dots\}$,
- výrokových logických spojek: $\neg$, $\wedge$,
- rovnosti: $=$,
- obecného kvantifikátoru $\forall$,
- operátoru znalosti K_i pro agenty $i, j, \dots$,

mimologických symbolů:

- Π je spočitatelná množina predikátových symbolů $(P_1, P_2, \dots)$,
- F je spočitatelná množina funkčních symbolů $(f_1, f_2, \dots)$,
- A je spočitatelná množina individuových konstant $(\alpha, \beta, \dots)$,

a pomocných symbolů:

- závorky: (,) a čárka ,

S každým predikátovým symbolem $P \in \Pi$ a funkčním symbolem $f \in F$ je asociováno přirozené číslo, které udává jeho aritu neboli četnost. P^n je n-ární predikátový symbol (tj. týká se n objektů), zatímco f^n je n-ární funkční symbol (tj. týká se n proměnných). Dále Π^n a F^n budou množiny predikátových a funkčních symbolů arity n.

4.4.2.1 Termy

Množina termů $\mathfrak{T}$ je nejmenší množina uzavřená vzhledem k následujícím pravidlům:

- každá proměnná a individuová konstanta je term: $V, A \subseteq \mathfrak{T}$,
- jestliže f^n s aritou n je funkční symbol a $\tau_1, ..., \tau_n$ jsou termy, pak $f(\tau_1, ..., \tau_n)$ je také term,
- nic jiného není term.

4.4.2.2 Formule

Jestliže P^n je n-místný predikátový symbol a $\tau_1, ..., \tau_n$ je n-tice termů, pak $P^n(\tau_1, ..., \tau_n)$ je atomická formule.

Množina $\mathcal{L}^*$ správně utvořených prvořádových epistemických formulí $\varphi, \psi...$ je nejmenší množina uzavřená vzhledem k následujícím podmínkám:

- každá atomická formule je správně utvořená formule,
- jestliže $\varphi, \psi \in \mathcal{L}^*$, pak $\neg\varphi, (\varphi \wedge \psi) \in \mathcal{L}^*$,
- jestliže $\tau_1, \tau_2 \in \mathcal{T}$, pak $(\tau_1 = \tau_2) \in \mathcal{L}^*$,
- jestliže $\varphi \in \mathcal{L}^*$ a $x \in V$, $\forall x(\varphi) \in \mathcal{L}^*$,
- jestliže $\varphi \in \mathcal{L}^*$, pak $K_i\varphi \in \mathcal{L}^*$ pro všechna $i \in A$,
- nic jiného není s.u.f.

Zavedeme opět klasické zkratky $(\varphi \vee \psi)$, $(\varphi \rightarrow \psi)$, $(\varphi \leftrightarrow \psi)$ pro $(\neg(\neg\varphi \wedge \neg\psi))$, $(\neg\varphi \vee \psi)$ a $((\varphi \rightarrow \psi) \rightarrow (\psi \rightarrow \varphi))$ v tomto pořadí. Dále budeme používat zkratku $\exists x(\varphi)$ pro $\neg\forall x(\neg\varphi)$. Rovněž budeme vynechávat vnější závorky vždy, když to bude možné. A vzhledem k tomu, že budeme pracovat jen s jedním agentem, tj. Alenkou, budeme vynechávat dolní index i a psát jen $K\varphi$ (množina A bude tedy singleton, tj. množina s jediným prvkem).

4.4.2.3 Sémantika SEL*

Stejně jako SEL i SEL* využívá Kripkeho sémantiku možných světů. Při přechodu z výrokového do predikátového kalkulu se ovšem musí odpovídajícím způsobem změnit struktura možných světů. Každému světu s

tak připadne množina individuí D (nazývaná též *doména*), kterou budou všechny světy sdílet, a interpretační funkce, která bude každému mimologickému výrazu připisovat intenzi, tj. funkci z možných světů do extenzí, přičemž extenzí mohou být právě individua nebo relace. Extenze každé formule bude tedy závislá na možném světě, v kterém bude ohodnocena.

Sémantika SEL* s jazykem $\mathcal{L}^*$ využívá tzv. *relační Kripkeho modely* $\mathcal{W}$, což je čtveřice tvaru $\langle D, S, \pi, \mathcal{R}\rangle$, kde:

- D je neprázdná množina individuí (doména); budeme předpokládat, že je konstantní napříč všemi světy, tj. že se jedná o společnou doménu,
- S je neprázdná množina možných světů,
- π je funkce, která každému světu přiřazuje relační model, přičemž relační model se skládá z domény D a přiřazení, které:
  - každému predikátovému symbolu P arity n přiřadí podmnožinu D^n množiny D,
  - každému funkčnímu symbolu f arity n přiřadí zobrazení z D^n do D,
  - každé individuové konstantě přiřadí prvek D,
- $\mathcal{R} \subseteq S \times S$ je relace ekvivalence.

Předpoklad tzv. *společné domény* zaručuje to, že všechny relační Kripkeho modely $\mathcal{W}$ sdílí stejnou doménu individuí. Jinými slovy, všechny možné světy musí obsahovat ta stejná individua. To rozhodně není neproblematický předpoklad a zde ho přijímáme jen kvůli zjednodušení.

Mějme relační Kripkeho model $\mathcal{W}$. Valuace v na $\mathcal{W}$ je funkce, která přiřazuje každé proměnné z V prvek domény D.

Nyní definujeme, co to znamená, že určitá formule φ je pravdivá ve světě s modelu $\mathcal{W}$ při valuaci v, což zapíšeme jako $(\mathcal{W}, s, v) \models \varphi$:

- $(\mathcal{W}, s, v) \models P(\tau_1, ..., \tau_n) \Leftrightarrow \left(v^{\pi(s)}(\tau_1), ..., v^{\pi(s)}(\tau_n)\right) \in P^{\pi(s)}$,

kde P je predikátový symbol arity n, $\tau_1, ..., \tau_n$ jsou termy a $v^{\pi(s)}$ je valuace termu vzhledem k relačnímu modelu přiřazenému světu s funkcí π. Dále definujeme identitu, negaci, konjunkci a obecný kvantifikátor:

- $(\mathcal{W}, s, v) \models (\tau_1 = \tau_2) \Leftrightarrow v^{\pi(s)}(\tau_1) = v^{\pi(s)}(\tau_2)$,
- $(\mathcal{W}, s, v) \models \neg\varphi \Leftrightarrow (\mathcal{W}, s, v) \not\models \varphi$,
- $(\mathcal{W}, s, v) \models \varphi \wedge \psi \Leftrightarrow (\mathcal{W}, s, v) \models \varphi$ a $(\mathcal{W}, s, v) \models \psi$,
- $(\mathcal{W}, s, v) \models \forall x(\varphi) \Leftrightarrow (\mathcal{W}, s, v_{(x/d)}) \models \varphi$ pro všechna d z D.

kde $v_{(x/d)}$ je taková valuace v', kdy $v'(y) = v(y)$ pro všechny proměnné y s případnou výjimkou x, přičemž valuace $v'(x) = d$. Na závěr jsme si nechali definici operátoru znalosti K. Ten je definován stejně jako v SEL:

- $(\mathcal{W}, s, v) \models K\varphi \Leftrightarrow (\mathcal{W}, t, v) \models \varphi$ pro všechna t taková, že $(s, t) \in \mathcal{R}$.

Formuli tvaru $K\varphi$ budeme chápat jako pravdivou v nějakém světě s tehdy, když formule $K\varphi$ bude pravdivá ve všech epistemických alternativách t ke světu s.

Nechť φ je formule, pak φ je *platná* v nějakém prvořádovém relačním Kripkeho modelu $\mathcal{W}$, jestliže pro všechny $s \in S$ a všechna přiřazení $v^{\pi(s)}$ platí $(\mathcal{W}, s, v) \models \varphi$. To zapíšeme jako $\mathcal{W} \models \varphi$. Dále řekneme, že formule φ je *platná*, jestliže $\mathcal{W} \models \varphi$ pro všechny prvořádné Kripkeho modely $\mathcal{W}$,

a to zapíšeme jako $\models \varphi$. A o formuli φ řekneme, že je *splnitelná*, jestliže je tu model $(\mathcal{W}, s, v)$ taková, že $\mathcal{W}$ je prvořádový Kripkeho model, kde $(\mathcal{W}, s, v) \models \varphi$.

Obdobně jako v SEL i zde můžeme relaci dosažitelnosti $\mathcal{R}$ udělit jisté podmínky. Nebudeme-li u relace $\mathcal{R}$ předpokládat žádné restrikce, dostaneme systém **K***, který je prvořádovým ekvivalentem systému **K** ze SEL. Systém **K*** je určen následujícími axiomy:

(FOL) Všechny tautologie predikátové logiky včetně axiomu rovnosti

(K) $K\varphi \wedge K(\varphi \to \psi) \to K\psi$

(BF) $\forall x\, K\varphi(x) \to K\, \forall x \varphi(x)$

(BC) $K\, \forall x \varphi(x) \to \forall x\, K\varphi(x)$

A odvozovací pravidla:

(MP) $\dfrac{\varphi \quad \varphi \to \psi}{\psi}$

(NEC) $\dfrac{\varphi}{K\varphi}$

(GEN) $\dfrac{\varphi}{\forall x(\varphi)}$

Systém **K*** můžeme dále rozšířit na **T***, **S4*** a **S5*** v tomto pořadí přidáváním následujících axiomů a korespondujících restrikcí na relaci $\mathcal{R}$.

(T*) $K\varphi \to \varphi$

(4*) $K\varphi \to KK\varphi$

(5*) $\neg K\varphi \to K\neg K\varphi$

Tyto prvořádové systémy **K*** až **S5*** jsou korektní a úplné vzhledem k odpovídajícím množinám modelů $\mathcal{W}$.

4.4.2.4 Vlastnosti SEL*

Modální logika s kvantifikátory s sebou obecně přináší řadu komplikací a ani SEL* není v tomto ohledu výjimkou.[41] Ukázalo se totiž, že v modální variantě predikátové logiky s identitou přestávají platit klasické extenzionální principy jako např. princip substitutivity identických entit, tj. možnost záměny koreferenčních termínů uvnitř modálních kontextů. S principem substitutivity identických entit jsme se setkali již dříve při vymezování extenzionálních systémů, nyní v podstatě jen zopakujeme to, co už jsme řekli, jen z pohledu SEL*.

Řekneme, že určitý kontext je *referenčně transparentní* k určitému termu τ_1 ve formuli φ právě tehdy, když τ_1 může být zaměněno v φ koreferenčním termínem τ_2 *salva veritate*, tj. platí:

$$\text{(PS)} \quad \forall\tau_1\forall\tau_2\Big((\tau_1 = \tau_2) \rightarrow \big(\varphi(\tau_1) \leftrightarrow \varphi(\tau_2)\big)\Big)$$

Jestliže (PS) neplatí, řekneme, že kontext formule φ je referenčně neprůhledný vůči termu τ_2.

Uvažme následující odvození, které si vypůjčuje postavy z povídky *Údolí strachu* od sira Artura C. Doyla:

$$\frac{\begin{array}{c}\text{McGinty ví, že Birdy Edwards je detektiv.}\\ \text{Birdy Edwards = McMurdo}\end{array}}{\text{McGinty ví, že McMurdo je detektiv.}}$$

[41] Jako jeden z prvních na ně upozornil [Quine, 1943].

Tento úsudek není v rámci modální logiky s kvantifikátory platný, byť (PS) by předpovídal jinak. Ale to je přesně to, čeho jsme chtěli celou dobu dosáhnout. Jinými slovy, je to právě toto selhání principu substitutivity identických entit v rámci SEL*, které nám umožňuje adekvátně analyzovat tento a jemu podobné epistemické úsudky.

Neplatnost (PS) lze přitom velmi jednoduše vysvětlit odkazem na tu skutečnost, že třebaže v tomto světě je Birdy Edwards identický s McMurdem, v jiných světech tomu tak být nemusí. Jinak řečeno, v jiných možných světech mohl A. C. Doyle napsat svůj příběh jinak a McMurdo mohl být identický např. s McGintym. Analogicky lze pak i vysvětlit selhání v případě Jitřenky a Večernice, Baracka Obamy a prezidenta USA atd. SEL* je schopna tyto rozdíly zohlednit právě z toho důvodu, že v porovnání se SEL pracuje s bohatšími pojmy možných světů. Přesněji, každá formule získává pravdivostní hodnotu v určitém světě s v závislosti na doméně individuí D a přiřazení, které je spojené s daným světem s. A to je právě to, co umožňuje, aby v jednom z možných světů platilo, že Birdy Edwards = McMurdo, ale v jiném světě se může jednat o dvě naprosto různá individua.[42]

Nyní přesuneme pozornost na další ze známých problému SEL*, a to kvantifikaci přes modální kontexty, resp. epistemickou interpretaci tzv. formule Barcanové.[43] Předpokládejme, že Alenka ví, že všechny muly jsou jalové. Tato její znalost může být ovšem zachycena pomocí dvou různých tvrzení:

(m) O každé mule Alenka ví, že je jalová.

[42] V této souvislosti stojí rovněž za zmínku Kripkeho teorie tzv. rigidních designátorů, tj. výrazů, které referují na stejné individuum ve všech možných světech, viz [Kripke, 1980].

[43] Srov. [Barcan, 1946].

(n) Alenka ví, že všechny muly jsou jalové.

O rozdílu mezi (m) a (n) se mluví jako o rozdílu mezi *de re* a *de dicto* znalostí v tomto pořadí. Čtení *de re* znamená to, že předmětem znalosti je samotný objekt, tj. v případě (m) muly. Naopak při čtení *de dicto* je předmětem znalosti celé tvrzení, tj. v případě (n) „Všechny muly jsou jalové“. Tvrzení (m) a (n) můžeme v SEL* zachytit následovně:

(m') $\forall x\, K\varphi(x)$

(n') $K\, \forall x \varphi(x)$

Jestliže spojíme (m') a (n') implikací, získáme formuli:

(BF) $\forall x\, K\varphi(x) \rightarrow K\, \forall x \varphi(x)$

což není nic jiného než dříve avizovaná formule Barcanové, resp. její epistemická interpretace. Slovní vyjádření: jestliže o každém x agent ví, že je φ, pak agent ví, že všechna x jsou φ. Zaměníme-li antecedent a konsekvent, dostaneme formuli:

(BC) $K\, \forall x \varphi(x) \rightarrow \forall x\, K\varphi(x)$

která je známá jako obrácená formule Barcanové a dozvídáme se z ní, že pokud agent ví, že všechna x jsou φ, pak o každém x agent ví, že je φ.

Jak je to s platností **(BF)** a **(BC)**? Lze je považovat za obecně epistemicky obhajitelné? Na první pohled by se mohlo zdát, že **(BC)** je obecně platná, ale není tomu tak. Předpoklad, že (n) je pravdivé, nezaručuje platnost (m), jak by naznačovala **(BC)**, neboť ve světě může existovat taková mula, o které Alenka neví, že je mulou. Tedy (n) si zachovává platnost, ale nikoli (m).

Podobně i v opačném směru. Z (m) nevyplývá nutně (n), neboť Alenka nemusí vědět, že jsou to všechny muly, které skutečně existují. Formulím **(BF)** a **(BC)** tedy nelze přiřknout obecnou platnost, nicméně jak **(BF)**, tak i **(BC)** jsou platné v naší SEL*, a to právě díky našemu zjednodušujícímu předpokladu společné domény. Je to totiž právě tento předpoklad, který zamezuje Alence v tom, aby uvažovala o nějakém světě s takovou mulou, o které neví, že je to mula.[44]

Shrnutí. Nyní, když už máme k dispozici SEL*, tj. standardní epistemickou logiku prvního řádu, jsme schopni analyzovat tvrzení, které odstartovalo celou tuto sekci, a to premisa (2a):

(2a) Alenka ví, že všichni lidé jsou smrtelní.

kterou nyní můžeme analyzovat jako:

(2a"') $K\,\forall x\,\big(\textsc{Člověk}(x) \rightarrow \textsc{Smrtelný}(x)\big)$

Čteme: Alenka ví, že pro všechna x platí, že pokud je x člověk, pak je x smrtelný. Nyní se konečně můžeme vrhnout na analýzu epistemických úsudků z druhé kapitoly.

4.4.3 Analýza epistemických úsudků

Epistemické úsudky (1) až (5) budeme analyzovat v rámci SEL*, a to konkrétně systémem **S5***. Přistupme k prvnímu úsudku:

[44] Problematika domén individuí hraje v rámci SEL* a modální predikátové logice obecně klíčovou roli, ale, jak už bylo řečeno, zde se spokojíme s předpokladem společné domény. Více o epistemické logice prvního řádu lze naleznout např. v [Gochet & Gribomont, 2006].

$$(1)\ \frac{\begin{array}{c}\text{Alenka ví, že prší.}\\ \text{Jestliže prší, pak není pravda, že neprší.}\end{array}}{\text{Alenka ví, že není pravda, že neprší.}}$$

který můžeme zachytit v SEL* následujícím způsobem:

$$(1''')\ \frac{\begin{array}{c}K\ \textsc{Prší}\\ \textsc{Prší} \to \neg\neg\textsc{Prší}\end{array}}{K\ \neg\neg\textsc{Prší}}$$

Tento úsudek je z hlediska SEL* stále platný. To je ovšem v rozporu s naším závěrem, ke kterému jsme dospěli v druhé kapitole. Přikročme nyní k úsudku (2):

$$(2)\ \frac{\begin{array}{c}\text{Alenka ví, že každý člověk je smrtelný.}\\ \text{Alenka ví, že Sokrates je člověk.}\end{array}}{\text{Alenka ví, že Sokrates je smrtelný.}}$$

který analyzujeme následovně:

$$(2''')\ \frac{\begin{array}{c}K\,\forall x\,\big(\textsc{Člověk}(x) \to \textsc{Smrtelný}(x)\big)\\ K\ \textsc{Člověk}(\textsc{Sokrates})\end{array}}{K\ \textsc{Smrtelný}(\textsc{Sokrates})}$$

I tento úsudek si v SEL* zachovává platnost, což se opět neshoduje s našimi výsledky. Pokračujme k úsudku (3):

$$(3)\ \frac{\begin{array}{c}\text{Alenka ví, že Barack Obama je Barack Obama.}\\ \text{Barack Obama je prezidentem USA.}\end{array}}{\text{Alenka ví, že Barack Obama je prezidentem USA.}}$$

který zachytíme jako:

$$(3''')\ \frac{\begin{array}{c}K\ (\textsc{Obama} = \textsc{Obama})\\ \textsc{Obama} = \textsc{PrezidentUSA}\end{array}}{\cancel{K\ (\textsc{Obama} = \textsc{PrezidentUSA})}}$$

Úsudek (3) již není v rámci SEL* platný, což je v souladu s naším závěrem z druhé kapitoly. Je zřejmé, že Barack Obama je Barack Obama, neboť se jedná o triviální tvrzení identity. Jinak řečeno, Barack Obama je ve všech možných světech vždy Barackem Obamou. Nicméně to, že Barack Obama je zrovna prezidentem USA, je empirické tvrzení, a tedy ne nutně pravdivé. Jsou představitelné takové možné světy, ve kterých je prezidentem USA někdo jiný, což je právě skutečnost, kterou dokáže SEL* zohlednit.[45] Postupme dále k úsudku (4):

(4) $$\frac{\begin{array}{c}\text{Alenka ví, že Narvik leží severně od Osla.}\\ \text{Alenka ví, že } x \text{ leží severně od } y \text{ právě tehdy, když } y \text{ leží jižně od } x.\end{array}}{\text{Alenka ví, že Oslo leží jižně od Narviku.}}$$

Analýza úsudku bude vypadat následovně:

(4”’) $$\frac{\begin{array}{c}K\ \big(\textsc{Severně}(\textsc{Narvik}, \textsc{Oslo})\big)\\ K\ \forall x \forall y\ \big(\textsc{Severně}(x, y) \leftrightarrow \textsc{Jižně}(y, x)\big)\end{array}}{K\ \big(\textsc{Jižně}(\textsc{Oslo}, \textsc{Narvik})\big)}$$

Úsudek (4) je v rámci SEL* rovněž platný, což se neshoduje s našimi dříve dosaženými výsledky. Pokud Alenka ví že Narvik leží severně od Osla a že nějaké x leží severně od y právě tehdy, když y leží jižně od x, pak musí vědět i to, že Oslo leží jižně od Narviku. Přesněji, k tomuto závěru nás zavazuje SEL*. Přistupme k poslednímu z našich úsudků (5):

[45] Analogickým způsobem lze vysvětlit komplikace, se kterými jsme se setkali při úsudku s Jitřenkou a Večernicí. Tvrzení „Jitřenka je Večernice“ je empirické zjištění, nejjasnější těleso na ranní obloze nemusí být nutně nejjasnější těleso na večerní obloze. Jinými slovy, lze si představit např. takové možné světy, ve kterých je Venuše Jitřenkou, ale ne Večernicí.

$$(5)\ \frac{\begin{array}{c}\text{Alenka ví, že } \big((p \wedge q) \supset p\big) \text{ je teorém.} \\ \big((p \wedge q) \supset p\big) \leftrightarrow \big((q \wedge p) \supset p\big)\end{array}}{\text{Alenka ví, že } \big((q \wedge p) \supset p\big) \text{ je teorém.}}$$

který zachytíme následovně:

$$(5''')\ \frac{\begin{array}{c}K\ \big((p \wedge q) \supset p\big) \\ \big((p \wedge q) \supset p\big) \leftrightarrow \big((q \wedge p) \supset p\big)\end{array}}{K\ \big((q \wedge p) \supset p\big)}$$

Toto odvození je v SEL* také platné, a to je opět v rozporu s našimi závěry z druhé kapitoly. Pokud Alenka ví, že $\big((p \wedge q) \supset p\big)$ (resp. že daná formule je teorém), rozhodně to ještě nemusí znamenat to, že ví, že i formule $\big((q \wedge p) \supset p\big)$ je teorém. Tyto formule jsou ale logicky ekvivalentní, a pokud Alenka ví, že $\big((p \wedge q) \supset p\big)$ je teorém, pak musí vědět i to, že $\big((q \wedge p) \supset p\big)$ je teorém.

Shrnutí. Intenzionální systém SEL* si při analýze epistemických úsudků vedl lépe než systémy extenzionální. S dosaženými výsledky avšak nemůžeme být spokojeni, neboť jednu správnou analýzu z pěti lze stěží považovat za úspěch. Jedna chybná analýza je dostatečným důvodem k zavržení systému explikace, o čtyřech nemluvě. Analýzy úsudků (1), (2), (4) a (5) jsme odmítli jako neadekvátní v důsledku jejich neintuitivních závěrů z hlediska platnosti. Tento problém, kdy nám samotný systém (v našem případě SEL*) vnucuje chybné, nepřijatelné či zkrátka nerealistické předpoklady o znalosti agenta, tj. že zná závěry, které znát nemusí, se obecně nazývá *problémem logické vševědoucnosti*.[46] Lze se tomu nějak vyhnout?

[46] Tento termín zavedl [Hintikka, 1975].

První možností je zcela odmítnout SEL* jako nevyhovující k výstavbě modelu znalosti. Avšak tento krok se zdá být poněkud unáhlený. Obyčejně nezavrhneme rovnou celý systém při prvním náznaku potíží, ale pokusíme se s nimi nějak vypořádat. V potaz je třeba vzít i to, že SEL* poskytuje mnoho výhod a umožňuje nám o znalosti uvažovat velmi intuitivním a pohodlným způsobem, tudíž se jí nebudeme chtít vzdát do té doby, dokud se to neprokáže jako skutečně nevyhnutelné. Stejně jako jsme se tedy pokusili dříve obhájit systémy extenzionální, nyní se pokusíme o to stejné v případě intenzionálních systémů, přesněji SEL*. Tím se dostáváme k druhé možnosti, a to pokusit se problém logické vševědoucnosti vyřešit v rámci intenzionálního systému SEL* s Kripkeho sémantikou.

Začneme tím, že blíže prozkoumáme samotnou logickou vševědoucnost. Řekli jsme si, že naše analýzy úsudků (1), (2), (4) a (5) se nezdařily v důsledku předpokladu logické vševědoucnosti agenta, který je v pozadí SEL*. Na jednotlivá selhání tak vlastně můžeme pohlížet jako symptomy mnohem závažnější poruchy, tj. právě logické vševědoucnosti. A pokud ji chceme odstranit, musíme ze všeho nejdříve řádně vyšetřit její příznaky, tj. blíže prozkoumat příčiny nesprávných závěrů (1), (2), (4) a (5). Začneme úsudkem (1), který jsme formálně zachytili následovně:

$$(1'') \quad \frac{\begin{array}{c} K\ \textsc{Prší} \\ \textsc{Prší} \rightarrow \neg\neg\textsc{Prší} \end{array}}{K\ \neg\neg\textsc{Prší}}$$

Ze sekce 4.3 o výstavbě SEL už víme, že jednou základních vlastností SEL* je to, že veškerá znalost je uzavřena vzhledem k logickému důsledku. Jinými slovy, agent musí znát všechny logické důsledky svých znalostí. Pokud tedy Alenka ví, že prší, pak si musí být schopna odvodit

i to, že není pravda, že neprší. Podle SEL* to Alenka zkrátka nemůže nevědět. Obecněji, zná-li agent φ, pak je φ pravdivé ve všech světech, které považuje za možné. A jestliže z φ logicky vyplývá ψ (tj. implikace $\varphi \to \psi$ je platná), pak ψ je také pravdivé ve všech světech, které agent považuje za možné. Z toho vyplývá, že agent musí znát také ψ. Tím se dostáváme k prvnímu principu logické vševědoucnosti. Formálněji:

(LO$_1$) jestliže $K\varphi$ a z φ vyplývá ψ, pak $K\psi$

Příčinou chybného vyhodnocení úsudku (1) je tedy **(LO)**$_1$. Pokračujme k úsudku (2):

$$(2'') \quad \frac{\begin{array}{c} K\,\forall x\,\big(\textsc{Člověk}(x) \to \textsc{Smrtelný}(x)\big) \\ K\ \textsc{Člověk}(\textsc{Sokrates}) \end{array}}{K\ \textsc{Smrtelný}(\textsc{Sokrates})}$$

Další základní vlastností SEL* byla uzavřenost znalosti vzhledem k materiální implikaci, resp. vzhledem k odvozovacímu pravidlu modus ponens. Obecně řečeno, pokud Alenka zná obě premisy, pak musí znát i závěr. Formálněji:

(**LO**$_2$) jestliže $K\varphi$ a $K(\varphi \to \psi)$, pak $K\psi$

Příčinou nesprávného vyhodnocení úsudku (2) je tedy **(LO)**$_2$.

Analýzu chybného vyhodnocení úsudku (4) přeskočíme, jelikož selhává z analogického důvodu jako úsudek (2) výše (vzpomeňme si, že ekvivalenci jsme definovali jako materiální implikaci oběma směry).

Přistupme tedy k poslednímu úsudku (5), který jsme zachytili jako:

$$(5''') \quad \frac{\begin{array}{c} K\ \big((p \wedge q) \supset p\big) \\ \big((p \wedge q) \supset p\big) \leftrightarrow \big((q \wedge p) \supset p\big) \end{array}}{K\ \big((q \wedge p) \supset p\big)}$$

Formule $\big((p \wedge q) \supset p\big)$ a $\big((q \wedge p) \supset p\big)$ jsou z logického (a v důsledku toho i sémantického) hlediska ekvivalentní, a proto je můžeme zaměnit.

Už z dřívějška víme, že možné světy se od sebe neliší logickými pravdami a význam určitého tvrzení (formule) je v intenzionálních systémech dán množinou možných světů, ve kterých je toto tvrzení pravdivé. Z toho tedy vyplývá, že pokud agent zná nějakou platnou logickou formuli (teorém), např. $\big((p \wedge q) \supset p\big)$, pak ví nejen to, že $\big((q \wedge p) \supset p\big)$ je rovněž teorém, ale zná i všechny další logické teorémy, což je zcela nepřijatelný závěr. Jak je to možné?

Jde o důsledek množinového pojetí propozic, přesněji toho, že v SEL* je význam pojímán jako propozice. A vzhledem k tomu, že všechna platná logická tvrzení označují stejnou propozici, znamená to, že jsou platná v těch stejných možných světech (tj. všech světech). Abychom to shrnuli, pokud agent ví, že nějaká formule φ je platná, a formule φ je logicky ekvivalentní s nějakou formulí ψ, pak agent musí vědět i to, že formule ψ je platná. Formálněji:

(**LO**$_3$) jestliže $K\varphi$ a formule $\varphi \leftrightarrow \psi$ je logicky platná, pak $K\psi$

Příčinou chybného vyhodnocení úsudku (5) je tedy **(LO)**$_3$.

Celkem tu máme tedy tři formy logické vševědoucnosti **(LO)**$_1$, **(LO)**$_2$ a **(LO)**$_3$ a obhajoba SEL* nebude spočívat v ničem jiném než v jejich odstranění. Je nutné zmínit, že tyto tři principy postihují stejným způsobem jak výrokovou, tak i predikátovou variantu našeho modelu znalosti, a proto budeme v rámci zjednodušení pracovat v následující kapitole jen se SEL.

Kapitola 5

Logická vševědoucnost a její řešení

V předchozí kapitole jsme si ukázali, že SEL umožňuje velmi snadným způsobem modelovat znalosti určitého agenta, nicméně, jak jsme také viděli, tento přístup k explikaci znalosti s pomocí Kripkeho sémantiky s sebou nese jisté problematické závazky. Pozorný čtenář si jistě už všiml, že formy logické vševědoucnosti **(LO$_1$)**, **(LO$_2$)**, **(LO$_3$)**, na které jsme narazili při analýze epistemických úsudků, nejsou vlastně nic jiného než dříve probírané základní vlastnosti našeho systému SEL, po řadě **(MON)**, **(K)** a **(CON)**. Tyto tři vlastnosti sdílí jednu společnou charakteristiku, a to, že se jedná o určité formy logického uzávěru znalostní báze agenta. V pozadí logické vševědoucnosti tedy stojí koncept logického důsledku. Pro připomenutí, z nějaké formule φ logicky vyplývá formule ψ, jestliže ψ platí vždy, když platí φ. To ovšem znamená, že v systému, kde je přítomné pravidlo necesitace **(NEC)**, agent musí znát i všechny teorémy daného systému, neboť ty jsou logickým důsledkem prázdné množiny

formulí. Tím se dostáváme k nové, v pořadí již čtvrté formě logické vševědoucnosti. Agent musí znát v rámci SEL všechny teorémy:

(**LO**$_4$) z formule φ odvoď $K\varphi$

Pokud lze určit pravdivost nějaké formule φ čistě na základě logiky (tj. je-li φ teorémem), pak to znamená, že formule φ je univerzálně epistemicky dostupná každému agentovi, neboť může být odvozena z prázdné množiny předpokladů. Jinak řečeno, každý agent v rámci SEL může vědět, že φ, jelikož φ je odvoditelná z libovolné znalostní báze.

Dohromady tedy dostáváme čtyři principy logické vševědoucnosti, které odpovídají čtyřem základním vlastnostem SEL:

- (**LO**$_1$) = **(MON)**,
- (**LO**$_2$) = **(K)**,
- (**LO**$_3$) = **(CON)**,
- (**LO**$_4$) = **(NEC)**.

To ovšem znamená, že zamezení logické vševědoucnosti nebude zdaleka tak jednoduché, jak se mohlo na první pohled zdát, neboť tyto čtyři vlastnosti (**LO**$_1$)–(**LO**$_4$), resp. **(MON)**, **(K)**, **(CON)** a **(NEC)**, jsou v SEL platné bez jakýchkoli předpokladů o relaci $\mathcal{R}$, tj. jsou pravdivé ve všech Kripkeho modelech $\mathbb{M}$. Jejich platnost tak vyplývá ze samotné podstaty Kripkeho sémantiky, která modeluje znalost jako nutnost. Jinými slovy, tyto čtyři principy nám vnucuje samotná kripkovská sémantika možných světů. Ať upravíme relaci $\mathcal{R}$ jakkoli, tyto vlastnosti si vždy zachovají platnost.[1]

[1] Proč nepracujeme jen s **(K)** a **(NEC)**, když jsme si řekli, že **(MON)** a **(CON)** z nich vyplývají? Důvod je ten, že někdy můžeme chtít např. jen omezit **(MON)**, ale zachovat **(K)**, proto je uvažujeme takto odděleně.

Je důležité dodat, že vedle (**LO**$_1$)–(**LO**$_4$) je tu ještě celá řada dalších forem logické vševědoucnosti, které v SEL platí. Můžeme tedy říci, že problém logické vševědoucnosti odkazuje na celou rodinu spřízněných logických uzávěrů, z nichž některé lze považovat za slabší, jiné za silnější:[2]

Zkratka	Formule	Název
(**LO**$_1$)	$\varphi \to \psi \Rightarrow K\varphi \to K\psi$	uzávěr vzhledem k logickému důsledku
(**LO**$_2$)	$K\varphi \wedge K(\varphi \to \psi) \to K\psi$	uzávěr vzhledem k materiální implikaci
(**LO**$_3$)	$\varphi \leftrightarrow \psi \Rightarrow K\varphi \leftrightarrow K\psi$	znalost ekvivalentních formulí
(**LO**$_4$)	$\varphi \Rightarrow K\varphi$	znalost platných formulí
(**LO**$_5$)	$(K\varphi \wedge K\psi) \to K(\varphi \wedge \psi)$	uzávěr vzhledem ke konjunkci
(**LO**$_6$)	$K\varphi \to K(\varphi \vee \psi)$	oslabení znalosti
(**LO**$_7$)	$\neg(K\varphi \wedge K\neg\varphi)$	konzistence znalosti
(**LO**$_8$)	$K(K\varphi \to \varphi)$	znalost znalosti
(**LO**$_9$)	$(K\varphi \wedge K\neg\varphi) \to K\psi$	znalost všech formulí

Zde se avšak budeme zabývat jen prvními čtyřmi, tj. (**LO**$_1$)–(**LO**$_4$), neboť ty vystupují jako nejzávažnější a kladou na agenta nejsilnější požadavky.

Můžeme tedy shrnout, že společně s Kripkeho sémantikou přijímáme následující omezení ohledně znalosti, kterou můžeme uvažovat, tj. znalost musí být uzavřena vzhledem k:

- logickému důsledku **(MON)**,
- materiální implikaci **(K)**,
- logické ekvivalenci **(CON)** a
- tautologiím **(NEC)**.

[2] Tabulka je s drobnými úpravami převzata z [Meyer & Hoek, 1995].

To jsou velmi problematické předpoklady. Uvědomíme-li si, že znalost zde pojímáme jako něco, k čemu se agent dostává usuzováním a dedukcí, z vlastností **(MON)**, **(K)**, **(CON)** a **(NEC)** vyplývá, že tu máme co dočinění s nevšedně nadaným, přímo dokonalým logikem. Stručně řečeno, agent je v rámci SEL modelován jako logicky vševědoucí, a to přímo jako aktuálně logicky vševědoucí, nikoli pouze potenciálně. To znamená, že agent dokáže nejen odvodit všechny důsledky svých znalostí, ale že je skutečně i odvodí. Jedná se vlastně o superlogika, který okamžitě zná všechny logické důsledky svých znalostí. A právě s takovým agentem pracuje SEL, tj. s logicky vševědoucím agentem.

Lidé ale nejsou logicky vševědoucí. Člověk se může naučit pravidla sudoku, a přesto nemusí být schopen vyluštit jedinou tabulku. Nechámeli stranou takové důvody jako únava, nepozornost, lenost, neochota, nezájem či nevůle připustit si jisté závěry, je hlavním zdrojem naší logické ignorance především to, že jsme při svých výpočtech a odvozeních závislý na určitých *zdrojích* (angl. resources) jako paměť či čas, které máme v obou případech k dispozici jen v omezeném množství. Nejsme zkrátka schopni všechno vypočítat a odvodit, protože na to nemáme dostatek času a paměti. Pro tuto skutečnost se užívá termín *omezenost zdrojů* (angl. resource-boundedness), jsme tedy tzv. *agenty s omezenými zdroji* (angl. resource-bounded agents).[3] Nesmíme ovšem zapomínat na to, že omezenost zdrojů není jedinou možnou příčinou logické nevševědoucnosti. Neschopnost dokončit konkrétní sudoku tabulku může mít mnohem prostší důvody jako např. pouhé neuvědomění si určitých souvislostí nebo prostě jen chybování a omyly.

[3] Tuto skutečnost zohledňují ve svých řešeních logické vševědoucnosti např. [Duc, 2001] a [Artemov & Kuznets, 2009].

Ať už je tomu jakkoli, závěr je stejný: nikdo nezná všechny logické důsledky všech svých znalostí, nikdo nezná všechny logické pravdy a není pravda, že pokud někdo ví, že platí nějaká formule φ, tak ví, že platí i všechny formule ψ, které jsou s φ ekvivalentní. Krátce řečeno, nikdo není logicky vševědoucí. Nicméně přesto toto je typ agenta, kterým jsme se doposud zabývali. Model znalosti SEL je tedy v mnoha ohledech značně nerealistický, a pokud chceme obhájit jeho smysluplnost — což rozhodně chceme — musíme se zbavit logické vševědoucnosti.

Postupovat můžeme dvěma způsoby. První možnost je to, že SEL necháme tak, jak je, ale pokusíme se obhájit její nerealistické důsledky v širším kontextu (sekce 5.1). Druhou možností je naopak to, že se pokusíme modifikovat samotný systém SEL tak, aby popisoval realističtější agenty, resp. takové agenty, kteří nejsou logicky vševědoucí (sekce 5.2).

5.1 Interpretace SEL v širším kontextu

Pokud SEL nemodeluje skutečné agenty, co tedy vlastně modeluje? Byl to pravděpodobně Lemmon, který jako první přišel s myšlenkou, že systémy jako SEL popisují tzv. ideálně racionálního agenta, neboť bylo jasné, že běžný člověk to být nemůže.[4] SEL tedy můžeme považovat za přijatelnou za předpokladu, že popisuje pouze ideálně racionální agenty. Nešlo vlastně o nic jiného než o problém obecné interpretace jednotlivých principů **(MON)**, **(K)**, **(CON)** a **(NEC)**, poněvadž nebylo zcela jasné, co mají vlastně představovat z širšího epistemologického hlediska. Lemmon na ně pohlížel právě jako na určité postuláty racionality, tj. nutné podmínky,

[4] Nutno ovšem poznamenat, že tohoto problému si byl vědom už sám Hintikka v [Hintikka, 1962].

které musí splňovat jeho racionální agent.[5] Takovéto kritérium racionality je ale příliš silné a vedlo by k tomu, že bychom ze SEL vykázali všechny agenty, kteří někdy pochybili v logice, tj. ty agenty, které za racionální běžně považujeme a pro které jsme SEL vůbec zavedli. Jinými slovy, nám nejde o epistemickou logiku ideálních agentů, ale o epistemickou logiku nás lidí. Racionalita a chybování se nijak nevylučují a na omylech není nic iracionálního.[6]

Jsou tu nějaké další návrhy jak chápat **(MON)**, **(K)**, **(CON)** a **(NEC)**? Vincent F. Hendricks navrhoval, abychom tyto vlastnosti pojímali jako určité *stupně neomylnosti* (angl. measures of infallibility). V rámci systému **K** jsou tak agenti např. neomylní v tom smyslu, že nikdy nemohou „opomenout" aplikaci pravidla modus ponens atd.[7] Lenzen se naopak kloní spíše k normativní interpretaci a SEL chápe jako normativní teorii racionálních postojů agenta.[8] Zajímavým způsobem navázal na tuto debatu Robert Stalnaker. Položil si obecnější otázku, co vlastně myslíme tím, když o SEL řekneme, že popisuje idealizovaného agenta, a rozlišil dva možné způsoby, jak chápat její idealizaci:

1. SEL popisuje normální agenty prostřednictvím idealizovaného pojmu znalosti, nebo

2. SEL popisuje idealizované agenty prostřednictvím běžného pojmu

[5] Srov. [Lemmon, 1967].

[6] Mnohem výhodnější je pohlížet na racionalitu ne jako na schopnost vyvozovat logické důsledky z toho, co víme, ale spíše jako na způsobilost uznat racionální úsudek, jeli nám předložen a náležitě vysvětlen. To odpovídá i naší běžné zkušenosti, neboť člověka nazýváme iracionálním spíše tehdy, když odmítá uznat závěry racionálních argumentů, ne když není schopen tyto argumenty sám vést.

[7] Srov. [Hendricks, 2006].

[8] SEL tedy nepopisuje to, jak pracujeme se znalostmi, ale jak bychom s nimi pracovat měli. Srov. [Lenzen, 2004].

znalosti.

Na jedné straně tak máme znalost v jakémsi speciálním technickém smyslu, na straně druhé agenty s neomezenou pamětí a nekonečnou výpočetní rychlostí. Je zřejmé, že Lemmonův konstrukt racionálního člověka by jednoznačně spadal do druhé kategorie.

Stalnaker dále uvádí celkem čtyři důvody, které mohou obhájit zavedení idealizace do určitého systému. Prvním z nich je lepší vysvětlení základního principu. V tomto smyslu je pak idealizovaná znalost (popř. agent) v podstatě analogií ideálního plynu či nakloněné, ideálně hladké roviny bez tření. Druhým důvodem je zjednodušení. Přestože jsou idealizace přísně vzato nepravdivé, přesto se je vyplatí v určitých kontextech užívat. Příkladem může být např. tvrzení, že paprsek světla vždy cestuje po přímce apod. Třetí způsob ospravedlnění idealizace je normativní. SEL nepopisuje to, jak uvažujeme, ale to, jak bychom uvažovat měli. Logická nevševědoucnost je tak nahlížena v podstatě jako jakýsi nežádoucí defekt, kterého se je třeba vyvarovat. Čtvrtý důvod, proč se zabývat idealizovanou SEL, je ten, že žádnou lepší epistemickou logiku zatím nemáme. Jinými slovy, SEL je sice značně idealizovaný model znalosti, ale je to stále ten nejlepší, co máme k dispozici.[9] Není třeba dodávat, že způsob ospravedlnění idealizace SEL se přímo odvíjí od toho, co od SEL vůbec očekáváme.

Určitá míra idealizace je samozřejmě vždy smysluplná, jak mj. ukázaly Stalnakerovy příklady, problém je ovšem ten, že idealizace prováděné SEL jsou příliš silné. Agent popisovaný SEL má velmi málo společného skutečným racionálním agentem. Je to spíše jen logická fikce,

[9] Srov. [Stalnaker, 1991], s. 426–430.

superlogik, ideální racionální člověk, teoretický konstrukt. Je tu ale ještě jeden způsob, jak vysvětlit nerealistické důsledky SEL, aniž bychom se museli uchylovat k silné idealizaci. K tomu ovšem budeme potřebovat dva nové pojmy, a to explicitní a implicitní znalost.

5.1.1 SEL jako model implicitní znalosti

Co je to explicitní a implicitní znalost? Vypomůžeme si následujícím příkladem. Alenka se dostala k rovnici:

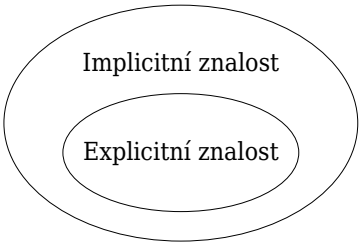
$$x^2 + 5x + 4 = 0.$$

Umí řešit kvadratické rovnice, takže ví, že kdyby chtěla, snadno by se mohla dopracovat k výsledku. Rozhodne se tedy příklad přeskočit a věnovat se raději obtížnějším příkladům. Alenka tedy sice nezná správnou odpověď (tj. hodnotu neznámé x), ale má dostatek informací k tomu, aby se k ní dopracovala. V tomto případě tedy představují kořeny rovnice výše Alenčinu *implicitní* znalost a řekneme, že zná výsledek implicitně. Nyní předpokládejme, že se situace opakuje, ovšem s tím rozdílem, že tentokrát Alenka rovnici vypočítá. Vyjde jí, že kořeny rovnice jsou (–1) a (–4). V této chvíli se jedná o *explicitní* znalost a o Alence můžeme říci, že zná výsledek explicitně.[10]

V čem je tedy rozdíl mezi implicitní a explicitní znalostí? Explicitní znalost je ta znalost, které si je agent vědom, kdežto implicitní znalost

[10] Distinkci implicitní/explicitní znalost zavedl Hector Levesque [Levesque, 1984]. Rescher místo implicitní znalosti používá termín *dostupná* znalost (angl. accessible knowledge), tj. taková znalost, ke které se agent může dostat, pokud dostatečně chytře nakládá s informacemi, které jsou mu explicitně přístupné. Srov. [Rescher, 2006]. Hintikka místo explicitní a implicitní znalosti používá termíny *aktivní* a *možná* (angl. virtual) znalost. Srov. [Hintikka, 1962].

je znalostí nevědomou. Implicitní znalost vlastně není znalost v pravém slova smyslu, ale je to spíše jen potenciální znalost, resp. soubor potenciálních znalostí, ke kterým by se mohl agent dostat, pokud by provedl odpovídající počet logických operací, odvození a výpočtů na své bázi znalostí. Explicitní znalosti jsou tak vlastně podmnožinou implicitních znalostí (viz obr. 5.1).



Obrázek 5.1: Implicitní a explicitní znalost

Je nutné si ale uvědomit, že ne vždy je tento proces převodu implicitních znalostí na explicitní tak snadný a bezproblémový jako tomu bylo v případě výše. Agent např. nemusí mít dostatek času, paměti a výpočetních dovedností k tomu, aby převedl svoje implicitní znalosti na znalosti explicitní.[11] V tomto ohledu byl příklad výše mírně idealizován, jelikož bral v potaz Alenčinu znalost řešení kvadratických rovnic. Ve skutečnosti by ovšem SEL připsala Alence implicitní znalost kořenů (–1) a (–4) i tehdy, když by o kvadratických rovnicích nikdy předtím neslyšela. Příklad výše byl tedy speciálním případem něčeho, co bychom mohli nazvat jako explicitní implicitní znalostí, tj. agent ví, že by mohl něco vědět, kdyby to vědět chtěl. Většinou ale agent řešení ani vědět nemůže, a to

[11] V této souvislosti se pak mluví o tzv. *proveditelné* (angl. feasible) znalosti, tj. takové znalosti, kterou je agent skutečně schopen „zexplicitnit".

i přesto, že by je vědět chtěl, a to právě třeba v důsledku nedostatku paměti a času (tzv. omezenost zdrojů), resp. v důsledku svých omezených výpočetních schopností.

Jak nám tato distinkce může pomoci odstranit logickou vševědoucnost, resp. nerealistické požadavky, které SEL klade na agenta? Řešení spočívá v tom, že formuli $K\varphi$ nebudeme číst jako „agent ví, že φ“, ale např. jako:

- „agent má dostatek informací k tomu, aby věděl φ“,
- „z agentových znalostí lze odvodit, že φ“,
- „agent ví implicitně, že platí φ“, „φ vyplývá z toho, co agent ví“ či
- „φ je potenciální znalost agenta“

  apod.

Logické vševědoucnosti se tedy můžeme v SEL zbavit tím, že poupravíme interpretaci jejich formulí.

Explicitní znalost pak zapíšeme K^e a formuli $K^e\varphi$ budeme číst jako „agent ví explicitně, že platí φ“ nebo „agent si je vědom toho, že φ je pravda“ atd. Pokud agent explicitně ví, že φ, znamená to, že dokáže zodpovědět otázku ohledně pravdivosti φ. Pokud bychom se zeptali té Alenky, která přeskočila řešení rovnice (tj. znala řešení pouze implicitně), zda jsou její kořeny čísla (–1) a (–4), nebyla by schopna pravdivě a upřímně odpovědět „Ano“ či „Ne“, ale spíše bychom se dočkali odpovědi jako „Nevím, je to možné“.

V této interpretaci logická vševědoucnost, resp. vlastnosti **(MON)**, **(K)**, **(CON)** a **(NEC)**, nadále nepředstavují problém. Jednoduše se řekne, že SEL nepopisuje to, co agent (explicitně) ví, ale pouze to, co by mohl

vědět s tím, co ví. Nejjednodušším způsobem, jak zamezit logické vševědoucnosti, je tedy reinterpretovat SEL tak, že nemodeluje to, co agent skutečně ví (explicitní znalost), ale pouze to, co by agent mohl vědět, resp. to, co je odvoditelné z formulí, které agent už ví (implicitní znalost). Tímto způsobem lze snadno ospravedlnit používání SEL, nicméně zaplatíme za to vysokou cenu.

Potíž spočívá v tom, že model znalosti předkládaný SEL je v této implicitní interpretaci zcela neinformativní, neboť se z něj nic nedozvídáme o agentově skutečné či dosažitelné znalosti. Jinými slovy, implicitní znalost je zcela bezcenná, nás totiž zajímá to, co agent skutečně ví (popř. co by si mohl skutečně odvodit), ne to, co by mohl snad potenciálně vědět, kdyby měl k dispozici neomezené zdroje. Jde nám tedy o explicitní znalosti, jelikož naše (vědomé) jednání a poznávání se řídí tím, co víme, ne tím, co bychom mohli teoreticky vědět s dostupnými informacemi a neomezenou pamětí a časem. Pokud řekneme, že se SEL zabývá výhradně explikací implicitní znalosti, zbavíme se sice logické vševědoucnosti, a tím ospravedlníme její smysluplnost, ale současně tím SEL vyškrtneme jako zcela neužitečnou z epistemického hlediska. S logickou vševědoucností se tedy pokusíme vypořádat jiným způsobem, a to dříve avizovanou modifikací SEL.

5.2 Modifikace SEL

Jak jsme viděli výše, systém SEL lze bez potíží používat, pokud chceme modelovat implicitní znalost nebo v případech, kdy nám nevadí, že modelujeme idealizovaný pojem znalosti, resp. znalost idealizovaného génia obdařeného logickou vševědoucností. Ale to by nám ve většině pří-

padů mělo vadit. Chtěli bychom naopak takovou epistemickou logiku, která by dokázala zohlednit jak neúplnost našich znalostí, tak i naši omezenou a chybující schopnost logické dedukce. Vyzkoušíme tedy druhou možnost a pokusíme se modifikovat samotnou SEL tak, aby vyšla vstříc i nedokonalým, neideálním, chybujícím a lenošným, přesto však racionálním agentům, tj. normálním lidem.

Řekli jsme si, že příčinou logické vševědoucnosti jsou principy **(MON)**, **(K)**, **(CON)** a **(NEC)**. Naše modifikace musí tedy směřovat k tomu, aby tyto čtyři principy pozbyly univerzální platnost. Pokud se nám to podaří, zbavíme se logické vševědoucnosti. Tento úkol se může na první pohled zdát jako neproveditelný, neboť, jak už víme, **(MON)**, **(K)**, **(CON)** a **(NEC)** jsou platné ve všech Kripkeho modelech a opírají se o pojem logického důsledku. Nicméně nesmíme zapomínat na to, že platnost a logický důsledek jsou relativní jak vzhledem k množině $\mathbb{M}$ modelů $\mathcal{M}$, tak i k pojetí samotných možných světů. To ovšem znamená, že pokud budeme chtít zamezit platnosti ($\mathbf{LO}_1$)–($\mathbf{LO}_4$), budeme muset vykročit za hranice SEL, resp. standardní logiky obecně. Jinými slovy, řešení problému logické vševědoucnosti musí nabídnout takový systém, ve kterém nebudou obecně platná pravidla **(MON)**, **(K)**, **(CON)** a **(NEC)**, ale který zároveň zachová naši intuitivní definici znalosti jako pravdivosti ve všech možných světech. Řešení bude tedy spočívat v zavedení určitých nestandardních logik. To by nemělo být překvapující, neboť, jak jsme si ukázali výše, je to právě standardní přístup, který logickou vševědoucnost zapříčiňuje.

Dohromady si představíme šest v literatuře pravděpodobně nejrozšířenějších způsobů, jak lze docílit odstranění či alespoň oslabení logické vševědoucnosti. Rovněž je třeba zdůraznit, že zde předložený přehled

bude převážně orientační. Z toho důvodu budou některé techničtější aspekty jednotlivých řešení vynechány, jiné mírně upraveny či zjednodušeny. Čtenáři, který by toužil po hlubším porozumění jednotlivých řešení včetně důkazů atd., doporučujeme vždy nahlédnout do pramenné literatury jednotlivých řešení.

První a v mnoha ohledech nejradikálnější řešení, které si představíme, mění celé pojetí možných světů zavedením tzv. *nemožných* světů (viz 5.2.1). Druhé řešení logické vševědoucnosti naopak využije *nekonzistentní* a *neúplné* světy (viz 5.2.2). Ve třetím řešení využijeme logiku s nestandardně definovanou negací (viz 5.2.3). Čtvrté řešení bude považovat pravdivost ve všech možných světech pouze za nutnou, nikoli však dostatečnou podmínku znalosti, přičemž dodatečným kritériem se stane *uvědomění* (viz 5.2.4). V pořadí páté řešení se pokusí vypořádat s logickou vševědoucností pomocí tzv. *předsudků* (viz 5.2.5) a nakonec se seznámíme s řešením, které definuje znalost jako pravdivost v podmnožině možných světů neboli *svazcích*, které jsou inspirovány Montague-Scottovou sémantikou (viz 5.2.6). Nyní už k samotným řešením.

5.2.1 Nemožné možné světy

Řekli jsme si, že možné světy se neliší logickými pravdami. Touto restrikcí vlastně zamezíme agentovi v tom, aby zaváděl do svých úvah nelogické, nesmyslné a iracionální možné světy. Nepochybně má smysl, aby agent uvažoval nad tím, kde si zapomenul klíče, ovšem přibírat do úvah i takové světy, v nichž si je zapomněl např. na třech místech současně, se jeví jako zcela zbytečné.

Toto omezení má ovšem dalekosáhlé a závažné důsledky. Tyto světy

totiž brání agentovi nejen v tom, aby se mýlil v logice, tj. neumožňují mu věřit logicky nekonzistentním tvrzením, ale neumožňují mu o nich ani uvažovat, což rozhodně odporuje naší každodenní zkušenosti. Důvodem je to, že pokud znalosti modelujeme v rámci možných světů, a možné světy chápeme pouze jako ty logicky možné, nezbude nám v systému místo pro logické nejistoty a omyly.

Uvažme např. situaci, kdy Alenka píše test z logiky a jejím úkolem je rozhodnout, které z následujících formulí jsou ekvivalentní formuli $\neg(p \wedge q)$:

a) $\neg(\neg p \vee q)$

b) $\neg p \vee \neg q$

c) $p \rightarrow \neg q$

Alenka vyloučila možnost a), ovšem není si jistá tím, zda je správně b) nebo c). Její současné přesvědčení tedy připouští dvě situace: první, ve které je se zadáním ekvivalentní b) a nikoli c), a druhou situaci, ve které je se zadáním ekvivalentní c) a nikoli b). Ani jednu z těchto situací ovšem není možné reprezentovat pomocí možných světů, jelikož vždy, když je se zadáním ekvivalentní b), je s ním ekvivalentní i c) a naopak.

Ačkoli by se tedy mohlo zdát, že množina možných světů, které agent bere v úvahu, jsou vlastní podmnožinou množiny všech objektivně (logicky) možných světů, není tomu vždy tak. Množina objektivně možných světů může být nejen větší, ale někdy i menší než množina možných světů daného agenta. Jinými slovy, agenti mohou považovat za možné i takově světy, které logicky vzato možné nejsou.

Jako příklad toho, že lidem skutečně nedělá problém věřit např. nekonzistentním tvrzením, si můžeme uvést tzv. *konjunkční omyl* (angl.

conjunction fallacy), který představili Daniel Kahneman a Amos Tversky.[12]

Představme si Lindu, je jí 31 let, je svobodná a velmi společenská. Má magisterský titul z filosofie a jako studentka se velmi zajímala o diskriminaci a sociální spravedlnost. Rovněž se podílela na několika protijaderných demonstracích. Která z následujících dvou možností je pravděpodobnější?

a) Linda pracuje v bance.

b) Linda pracuje v bance a je aktivní členkou feministického hnutí.

Celkem 89 % dotázaných zvolilo možnost b). To je ovšem špatně, neboť pracovnice v bance, které jsou i aktivními feministkami, tvoří vlastní podmnožinu množiny všech pracovnic v bance, tudíž b) nemůže mít větší pravděpodobnost než a).

Je zřejmé, že naše znalosti (resp. přesvědčení) nemusí být zdaleka vždy konzistentní. I racionální agenti často operují v rámci nekonzistentního obrazu světa. A právě to je jeden z důvodů pro zavedení *nemožných světů* (angl. impossible worlds). Nemožné světy, jak název napovídá, jsou zkrátka ty světy, ve kterých je svým způsobem možné i to nemožné.

Z toho vyplývá, že v nemožných světech nemusí platit žádné logické zákony. Může v nich tak např. platit kontradikce, z příkladu výše může být b) pravděpodobnější než a) atp. Hypotéza nemožných světů se může na první pohled zdát jako absurdní, ale nesmíme zapomínat na to, že nám jde o epistemicky možné světy nikoli objektivně (logicky) možné světy. Není nic nepředstavitelného na tom, že se agent může domnívat,

[12] Srov. [Tversky & Kahneman, 1983].

že formule (resp. formulové schéma) $\varphi \wedge \neg\varphi$ je platná. A právě tuto intuici nám umožňují zachytit nemožné světy.

Můžeme shrnout, že nemožné světy jsou takové světy, ve kterých je možné úplně všechno, resp. vše, co si agent usmyslí (a tedy i neplatnost logických pravd). V tomto ohledu jsou nemožné světy vlastně jen výplodem fantazie daného agenta. Jak poznamenává Vincent F. Hendricks, jsou to světy ještě horší než ty s descartovským démonem, neboť ty byly alespoň logicky možné.[13] Nemožné světy jsou tedy pouze zdánlivě možné světy, tj. světy, které se agentovi pouze jeví jako možné, ale ve skutečnosti možné nejsou. Pomocí nich vlastně zavadíme do SEL lidskou omezenost, slabost, lenost, omylnost, hloupost atd. A to je přesně to, co potřebuje k zamezení logické vševědoucnosti, jelikož tyto světy umožňují agentovi nevědět ani ty nejzákladnější logické pravdy.

Po technické stránce se myšlenka nemožných světů opírá o Kripkeho teorii *nenormálních světů* (angl. non-normal worlds),[14] přičemž jejich užitečnosti při řešení problematiky propozičních postojů si pravděpodobně poprvé všiml Max Cresswell, který je označoval jako *neklasické světy* (angl. non-classical worlds), a jejich přímé aplikovatelnosti v epistemické logice pak Hintikka,[15] který je užíval pod již známým jménem nemožné světy. Formálně je pak zachytil jeho krajan Veikko Rantala.[16]

Rantalův návrh k překonání logické vševědoucnosti je založen na již zmíněné Kripkeho teorii nenormálních světů. V tomto přístupu je množina (normálních) možných světů rozšířena o množinu S^* nemožných světů, které se mohou právě chovat nelogicky.

[13] Srov. [Hendricks, 2006], s. 100.
[14] Srov. [Kripke, 1965].
[15] Srov. [Hintikka & Hintikka, 1989].
[16] Srov. [Rantala, 1982].

Nechť $\mathcal{L}$ je jazyk výrokové modální logiky a $\mathcal{P}$ neprázdná množina atomických formulí, pak Rantalův model je model $\mathcal{M}$ tvaru $\langle S, S^*, \pi, \pi^*, \mathcal{R}, \rangle$, kde:

- S je neprázdná množina možných světů,
- S^* je množina nemožných světů,
- π je funkce, která přiřazuje pravdivostní hodnotu atomickým formulím v možných světech S, tj. $\pi : S \mapsto (\mathcal{P} \mapsto \{\texttt{true}, \texttt{false}\})$,
- π^* je funkce, která přiřazuje pravdivostní hodnotu libovolným formulím v nemožných světech S^*, tj. $\pi^* : S^* \mapsto (\mathcal{L} \mapsto \{\texttt{true}, \texttt{false}\})$,
- $\mathcal{R}$ je relace dosažitelnosti taková, že $\mathcal{R} \subseteq (S \cup S^*) \times (S \cup S^*)$, tj. spojuje možné nebo nemožné světy s dosažitelnými možnými nebo nemožnými světy.

Formule jsou v možných světech S interpretovány stejně jako v Kripkeho modelech, a to včetně modálního operátoru K, avšak interpretace formulí v nemožných světech S^* je zcela libovolná. Přesněji, se složenými (tj. ne atomickými) formulemi je v nemožných světech nakládáno stejně jako s atomickými formulemi, a pravdivostní hodnotu jim tedy přiřazuje přímo funkce π^*:

- $(\mathcal{M}, s^*) \models \varphi$ právě tehdy, když $\pi^*(s^*)(\varphi) = \texttt{true}$.

To znamená, že interpretace složených formulí je zcela nahodilá, nezávislá na interpretaci jejích částí, což nám ve výsledku umožňuje např. formuli $\varphi \wedge \neg\varphi$ přidělit pravdivostní hodnota `true` v nějakém nemožném světě $s^* \in S^*$.

V důsledku naprosté volnosti při interpretaci formulí v nemožných světech díky π^* je sémantika nemožných světů silným nástrojem proti logické vševědoucnosti. Agent nemusí disponovat ani jediným obecně platným logickým principem, a snadno se tak můžeme vyhnout principům logické vševědoucnosti (**LO**$_1$) až (**LO**$_4$).[17] Uvažme např. (**LO**$_2$). Mějme model $\mathcal{M} = \langle S, S^*, \pi, \pi^*, \mathcal{R}, \rangle$, kde $S = \{s\}$, $S^* = \{s^*\}$ a dále:

- $\pi(s)(p) = \texttt{true}$,
- $\pi(s)(q) = \texttt{true}$,
- $(\mathcal{M}, s) \models p \to q$
- $\pi^*(s^*)(p) = \texttt{true}$,
- $\pi^*(s^*)(p \to q) = \texttt{true}$,
- $\pi^*(s^*)(q) = \texttt{false}$.

přičemž $\mathcal{R}(s, s^*)$. V takovém světě s platí jak Kp, tak i $K(p \to q)$ (protože formule p a $p \to q$ jsou pravdivé ve všech dosažitelných světech, tj. s a s^*), ale už neplatí Kq (protože formule q není pravdivá v dosažitelném světě s^*). Jinými slovy, o takovémto světě s můžeme říci, že v něm neplatí (**LO**$_2$), resp. $K\varphi \wedge K(\varphi \to \psi) \to K\psi$.

S problémem logické vševědoucnosti se tak lze vypořádat velmi snadno a rychle. Slabou stránkou tohoto řešení je ovšem to, že je poněkud arbitrární. Každá formule může být v této sémantice prohlášena za platnou nebo neplatnou v důsledku libovolně zvolených nemožných světů. Vhodnou volbou relace dosažitelnosti $\mathcal{R}$, resp. nemožného světa s^* a příslušné funkce π^*, není problém popřít ani platnost těch nejjednodušších tautologií jako $\varphi \to \varphi$. Na tom samozřejmě není nic špatného, ale problém spočívá v tom, že sémantiku nemožných světů lze tak stěží nazvat *logickou* sémantikou v přísném slova smyslu, v důsledku čehož se náš systém stává (alespoň z logického hlediska) dosti nezajímavým.

[17] Srov. [Meyer & Hoek, 1995].

5.2.2 Explicitní znalost

Plodným způsobem se s logickou vševědoucností snažil vypořádat Levesque.[18] Ve svém řešení logické vševědoucnosti zohledňuje rozdíl mezi explicitní a implicitní znalostí (resp. přesvědčením), ale pokračuje ještě o krok dále k jeho formálnímu zachycení. Jen pro připomenutí, myšlenka v pozadí je ta, že agent si sice musí být vědom všech svých znalostí (jinak by to nebyly znalosti), nicméně nemusí si uvědomovat všechny jejich logické důsledky, tj. implicitní znalosti. Podstata tohoto řešení logické vševědoucnosti, které se vztahuje pouze na explicitní znalost, spočívá ve zmírnění podmínky, že světy musí být úplné a konzistentní. Po technické stránce je toho dosaženo rozlišením pravdivostních a nepravdivostních podmínek pro formule. Přesněji, zavedeme dvě relace splnitelnosti $\models$, přičemž první z nich bude určovat pravdivost formulí, druhá jejich nepravdivost. Budeme tedy pracovat se dvěma relacemi $\models$ místo jedné.

Levesque zakládá svoji sémantiku na tzv. *situacích*, které můžeme chápat jako neklasické možné světy; neklasické v tom smyslu, že na rozdíl od klasických (normálních, standardních) možných světů, ty Levesquovy nemusejí být úplné. Myšlenka v pozadí je ta, že situace jsou takové „mikro-světy“, méně obsáhlé než regulérní světy, a tudíž nemusí obsahovat pravdivostní ohodnocení všech tvrzení, tj. některá v nich mohou být pravdivá, některá nepravdivá a jiná v nich nemusí mít pravdivostní hodnotu vůbec. Jedná se tedy o parciální (částečné) možné světy. Zde se ovšem Levesque nezastavuje a povoluje i nekonzistentní světy, tj. situace, se kterými není kompatibilní žádný možný svět. Jedná se tedy o situace, které podporují současně jak pravdu, tak i nepravdu něja-

[18] Srov. [Levesque, 1984].

kého tvrzení. Pokud si to shrneme, v Levesquových světech může být určitá formule pravdivá, nepravdivá, nemít pravdivostní hodnotu (neúplné světy) anebo nabývat obou pravdivostních hodnot současně (nekonzistentní světy), přičemž svět nazveme klasickým, není-li ani nekonzistentní, ani neúplný. Dochází tedy opět ke značnému odklonu od SEL, podobně jako v případě nemožných světů.

Nyní k formálnímu představení jeho řešení. Pro zjednodušení vynecháme implicitní znalost, neboť, jak už bylo výše zmíněno, ta si ponechává všechny charakteristické rysy logické vševědoucnosti. Jinými slovy, implicitní znalost je stále uzavřena vzhledem k logické a materiální implikaci, ekvivalenci a vzhledem k tautologiím. To je důsledkem toho, že implicitní znalost je na rozdíl od explicitní znalosti definována pomocí klasických světů, nikoli těch nekonzistentních a neúplných.

Nechť $\mathcal{L}^e$ je jazyk výrokové modální logiky rozšířený o operátor explicitní znalosti K^e a $\mathcal{P}$ neprázdná množina atomických formulí, pak Levesquův model $\mathcal{M}$ má tvar $\langle S, \mathcal{B}, \pi_{\texttt{true}}, \pi_{\texttt{false}} \rangle$, kde:

- S je neprázdná množina možných světů,
- $\mathcal{B} \subseteq S$ je množina světů, které agent explicitně považuje za možné,
- $\pi_{\texttt{true}}, \pi_{\texttt{false}}$ jsou funkce, které přiřazují pravdivostní hodnotu atomickým formulím v možných světech S, tj.
  $\pi_{\texttt{true}}, \pi_{\texttt{false}} : S \mapsto (\mathcal{P} \mapsto \{\texttt{true}, \texttt{false}\})$; $\pi_{\texttt{true}}(p)$ tedy obsahuje všechny světy, ve kterých je p pravdivé, a $\pi_{\texttt{false}}(p)$ všechny ty světy, ve kterých je p nepravdivé. Svět s může být nekonzistentní i neúplný. Relace $\models_{\texttt{t}}, \models_{\texttt{f}}$ mezi světy a formulemi jsou definovány následovně:
  - $(\mathcal{M}, s) \models_{\texttt{t}} p \Leftrightarrow \pi_{\texttt{true}}(s)(p) = \texttt{true}$,

- $(\mathcal{M}, s) \models_{\mathtt{f}} p \Leftrightarrow \pi_{\mathtt{false}}(s)(p) = \mathtt{true}$,
- $(\mathcal{M}, s) \models_{\mathtt{t}} (\varphi \vee \psi) \Leftrightarrow (\mathcal{M}, s) \models_{\mathtt{t}} \varphi$ nebo $(\mathcal{M}, s) \models_{\mathtt{t}} \psi$,
- $(\mathcal{M}, s) \models_{\mathtt{f}} (\varphi \vee \psi) \Leftrightarrow (\mathcal{M}, s) \models_{\mathtt{f}} \varphi$ a $(\mathcal{M}, s) \models_{\mathtt{f}} \psi$,
- $(\mathcal{M}, s) \models_{\mathtt{t}} (\varphi \wedge \psi) \Leftrightarrow (\mathcal{M}, s) \models_{\mathtt{t}} \varphi$ a $(\mathcal{M}, s) \models_{\mathtt{t}} \psi$,
- $(\mathcal{M}, s) \models_{\mathtt{f}} (\varphi \wedge \psi) \Leftrightarrow (\mathcal{M}, s) \models_{\mathtt{f}} \varphi$ nebo $(\mathcal{M}, s) \models_{\mathtt{f}} \psi$,
- $(\mathcal{M}, s) \models_{\mathtt{t}} \neg\varphi \Leftrightarrow (\mathcal{M}, s) \models_{\mathtt{f}} \varphi$,
- $(\mathcal{M}, s) \models_{\mathtt{f}} \neg\varphi \Leftrightarrow (\mathcal{M}, s) \models_{\mathtt{t}} \varphi$.

To znamená, že $(\mathcal{M}, s) \models_{\mathtt{t}} p$ označuje tu skutečnost, že p je pravdivé v s pro určité $p \in \mathcal{P}$, zatímco $(\mathcal{M}, s) \models_{\mathtt{f}} p$ znamená, že p je nepravdivé v s. Operátor znalosti K^e je pak definován takto:

- $(\mathcal{M}, s) \models_{\mathtt{t}} K^e\varphi \Leftrightarrow (\mathcal{M}, t) \models_{\mathtt{t}} \varphi$ pro všechna $t \in \mathcal{B}$,
- $(\mathcal{M}, s) \models_{\mathtt{f}} K^e\varphi \Leftrightarrow (\mathcal{M}, s) \not\models_{\mathtt{t}} K^e\varphi$.

Je zřejmé, že nekonzistentní světy hrají v podstatě stejnou úlohu jako nemožné světy v předcházejícím řešení. Rovněž si všimněme, že v tomto řešení je zcela vynechána relace dosažitelnosti $\mathcal{R}$, jejíž úlohu v podstatě supluje $\mathcal{B}$, což je další značný odklon od SEL.

I tento přístup umí řešit všechny čtyři formy logické vševědoucnosti, tj. (**LO**$_1$), (**LO**$_2$), (**LO**$_3$) a (**LO**$_4$).[19] Uvažme znovu (**LO**$_2$). Chceme ukázat, že neplatí $K\varphi \wedge K(\varphi \rightarrow \psi) \rightarrow K\psi$. Mějme model $\mathcal{M} = \langle S, \mathcal{B}, \pi_{\mathtt{true}}, \pi_{\mathtt{false}} \rangle$, kde $\mathcal{B} = \{t\}$. Dále předpokládejme, že:

- $\pi_{\mathtt{true}}(t)(p) = \mathtt{true}$,
- $\pi_{\mathtt{false}}(t)(p) = \mathtt{true}$ a

[19] Srov. [Meyer & Hoek, 1995].

- $\pi_{\texttt{true}}(t)(q) = \texttt{false}$.

To znamená, že:

- $(\mathcal{M}, s) \models_{\texttt{t}} K^e p$ a
- $(\mathcal{M}, s) \models_{\texttt{t}} K^e(p \to q)$, ale ne
- $(\mathcal{M}, s) \models_{\texttt{t}} K^e q$.

Tudíž explicitní znalost není uzavřena vzhledem k materiální implikaci. Je to tedy právě přítomnost nějakého nekonzistentního světa, která nám umožňuje popřít platnost ($\mathbf{LO}_1$).

Agent je dále také schopen držet nekonzistentní přesvědčení (tj. $Kp \wedge K\neg p$), ale jen v tom případě, že všechny možné světy, o kterých agent uvažuje, jsou rovněž nekonzistentní. To dává smysl i intuitivně, neboť kdybychom měli na výběr mezi konzistentním a nekonzistentním vysvětlením, těžko bychom volili to nekonzistentní. Je zřejmé, že tyto nekonzistentní a neúplné světy nejsou vlastně nic jiného než umírněnější forma nemožných světů představených výše.

Levesquův systém má ovšem hned několik nevýhod. V jeho logice explicitní znalosti se nemohou vyskytovat vnořené operátory, nejsou tedy přípustné formule jako $K^e K^e \varphi$. Z toho ovšem vyplývá, že jeho systém není schopen modelovat systémy s více agenty, tj. je omezen pouze na monoagentní systémy. Druhým nežádoucím důsledkem je to, že agent není schopen žádné formy introspekce, tj. znalosti o znalosti. Pokud jde o nekonzistentní světy, Levesque si byl dobře vědom jejich problematického statusu. Dle něj se jednalo o světy, které jsou sice (subjektivně) představitelné, ale nikoli (objektivně) uskutečnitelné. Nekonzistence tak vlastně pramení z epistemického stavu agenta, nikoli z nekonzistence

světa jako takového a předpoklad, že se agent může dostat k nekonzistentním informacím např. ze dvou z různých zdrojů, se jeví naprosto v pořádku.

5.2.3 Nestandardní logiky

Fagin et al. přišli s řešením logické vševědoucnosti využívající nestandardní výrokovou logiku (NVL),[20] která podobně jako Levesquova logika explicitní znalosti rozděluje sémantiku na dvě části. Přesněji, odděluje sémantiku formulí od sémantiky jejich negace. Jejich řešení je tak vlastně alternativou k Levesquovi, která umožňuje dosáhnout podobných výsledků, ovšem bez nutnosti postulovat nekonzistentní světy.

Nyní k formálnímu zachycení. Pro každý svět $s \in S$ budeme mít zrcadlový svět $\overline{s}$ takový, že formule $\neg\varphi$ je pravdivá ve světě s právě tehdy, když formule φ je nepravdivá ve světě $\overline{s}$. Místo toho, aby tedy formule $\neg\varphi$ platila ve světě s právě tehdy, když formule φ neplatí ve světě s, budeme tedy vyžadovat, aby $\neg\varphi$ platila ve světě s právě tehdy, když φ neplatí ve světě $\overline{s}$. Svět $\overline{s}$ tak můžeme chápat jako doplněk světa s, tj. $\overline{\overline{s}} = s$, přičemž svět je standardní, jestliže platí $s = \overline{s}$. Cílem je vlastně dosáhnout toho, aby pravdivostní hodnota $\neg\varphi$ nezávisela na pravdivostní hodnotě φ. Stručně řečeno, svět $\overline{s}$ slouží k interpretaci formulí s negací.

Nechť $\mathcal{L}^{-}$ je jazyk výrokové modální logiky a $\mathcal{P}$ neprázdná množina atomických formulí, pak NVL model je model $\mathcal{M}$ tvaru $\langle S, \pi, \mathcal{R}, ^{-} \rangle$, kde:

- S je neprázdná množina možných světů,
- π je funkce, která přiřazuje pravdivostní hodnotu atomickým formulím v možných světech, tj. $\pi : S \mapsto (\mathcal{P} \mapsto \{\texttt{true}, \texttt{false}\})$,

[20] Srov. [Fagin et al., 1995].

- $\mathcal{R}$ je relace dosažitelnosti (ekvivalence),
- $^-$ je unární funkce, jejíž definiční obor a obor hodnot se rovná S. Jinými slovy, $^-$ je funkce z možných světů S do sebe sama.

Sémantika jazyka $\mathcal{L}^-$ zůstává stejná jako v běžném Kripkeho modelu, pouze je třeba upravit definici pro negaci:

- $(\mathcal{M}, s) \models \neg\varphi \Leftrightarrow (\mathcal{M}, \overline{s}) \not\models \varphi$.

Slovně: formule $\neg\varphi$ je platná ve světě s právě když, když formule φ není platná ve světě $\overline{s}$. Od klasické výrokové logiky se zde tedy odchylujeme definicí $\neg\varphi$, neboť $\neg\varphi$ je pravdivé v $(\mathcal{M}, s)$, jestliže φ je nepravdivé vzhledem k druhému pravdivostnímu přiřazení v $\mathcal{M}$, tj. jestliže je nepravdivé v $(\mathcal{M}, \overline{s})$.

Tento přístup řeší **(MON)** a **(K)**. Pravidlo generalizace znalosti sice stále platí, ale NVL v důsledku naší nestandardní definice negace neobsahuje žádné tautologie, tudíž ani **(NEC)** nepředstavuje problém.[21] Platnost si ovšem ponechá **(CON)**. Tyto nestandardní Kripkeho modely jsou v mnoho ohledech ekvivalentní s těmi Levesqueho, ovšem vyhýbají se některým jejich neduhům jako např. zavádění nekonzistentních světů či nemožnosti výskytu vnořených operátorů.

5.2.4 Uvědomění

Ronald Fagin a Joseph Y. Halpern vyzkoušeli při řešení logické vševědoucnosti syntaktičtější přístup opírající se o pojem *uvědomění* (angl.

[21] Srov. [Fagin et al., 1995].

awareness).[22] Jak nám může koncept uvědomění pomoci při řešení logické vševědoucnosti? Základní myšlenka v pozadí je ta, že nestačí jen zjistit, zda ta či ona znalost vyplývá z agentovy báze znalostí, ale musíme zkontrolovat i to, zda si jí agent uvědomil. Úkolem uvědomění je tedy určovat formule, které si agent skutečně uvědomuje a které nikoli, tj. které mu pouze připisuje SEL v podobě implicitní znalosti. Zavazuje-li nás systém k nějakým nežádoucím závěrům ohledně znalosti agenta, jednoduše tyto nepohodlné znalosti vykážeme z jeho uvědomění. Tento přístup tedy zavádí dodatečnou podmínku pro připisování znalosti, a to že znalost musí být uvědomělá.

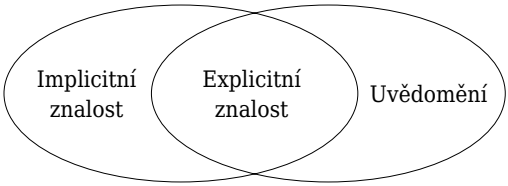
Kripkeho model tentokrát rozšíříme o funkci $\mathcal{A}$, která bude filtrovat explicitní znalosti z množiny implicitních znalostí. Jinými slovy, funkce $\mathcal{A}$ nám poslouží jako určité síto, pomocí kterého budeme určovat znalosti, jež si agent uvědomuje. Skutečnost, že si agent uvědomil nějakou formuli φ, zapíšeme jako $A\varphi$. Uvědomění je tedy syntaktický koncept, jelikož operuje přímo na konkrétních neinterpretovaných formulích.[23] Na druhou stranu, stále je tu i logická sémantika, která se stará o platnost formulí bez operátoru A.[24] Asi není potřeba dodávat, že uvědomění neimplikuje znalost, tj. agent si může uvědomit formuli φ, aniž by věděl, že φ. Intuitivně, explicitní znalost je ta implicitní znalost, kterou si agent uvědomil (viz obr. 5.2).

Tento systém je známý jako logika obecného uvědomění (angl. general awareness logic neboli GAL). Proč obecného? Protože funkci uvědo-

[22] Srov. [Fagin & Halpern, 1988].

[23] Více o řešení logické vševědoucnosti z pohledu syntaktického pojetí znalosti v Dodatku B.

[24] Čistě sémantický a plně rekurzivní model znalosti by byl samozřejmě vhodnější, ale, jak se domnívá např. Elias Thijsse, ten je nedosažitelný vzhledem k psychologické povaze uvědomění jako takového. Srov. [Thijsse, 1993].



Obrázek 5.2: Vztah znalosti a uvědomění

mění lze aplikovat na libovolné formule, tj. jak na atomické, tak i složené.

Nechť $\mathcal{L}^{eA}$ je jazyk výrokové modální logiky obohacený o operátor uvědomění A a explicitní operátor znalosti K^e a $\mathcal{P}$ neprázdná množina atomických formulí, pak GAL model $\mathcal{M}$ má tvar $\langle S, \pi, \mathcal{R}, \mathcal{A}\rangle$, kde:

- S je neprázdná množina možných světů,
- π je funkce, která přiřazuje pravdivostní hodnotu atomickým formulím v možných světech, tj. $\pi : S \mapsto (\mathcal{P} \mapsto \{\texttt{true}, \texttt{false}\})$,
- $\mathcal{R}$ je relace dosažitelnosti (ekvivalence),
- $\mathcal{A}$ je funkce, která přiřazuje agentovi v každém světě $s \in S$ množinu formulí, které si uvědomuje, tj. $\mathcal{A} : S \mapsto \wp(\mathcal{L}^{EA})$.

Sémantika pro atomické formule, negaci a konjunkci zůstává stejná jako v SEL, jen je třeba doplnit definici pro nový operátor A:

- $(\mathcal{M}, s) \models A\varphi \Leftrightarrow \varphi \in \mathcal{A}(s)$.

Slovně: agent si uvědomuje formuli φ právě tehdy, když je formule φ prvkem množiny, kterou agentovi přiřazuje funkce $\mathcal{A}$. Explicitní znalost definujeme za pomoci uvědomění:

- $(\mathcal{M}, s) \models K^e\varphi \Leftrightarrow \varphi \in \mathcal{A}(s)$ a $(\mathcal{M}, t) \models \varphi$ pro všechna t taková, že $(s, t) \in \mathcal{R}$.

Dostaneme tedy, že $K^e\varphi \leftrightarrow (A\varphi \wedge K\varphi)$, tj. explicitní znalost je implicitní znalost, které si je agent vědom, což odpovídá i intuitivnímu významu.

Stejně jako předchozí řešení i toto je schopno zamezit logické vševědoucnosti ve všech případech (**LO**$_1$)–(**LO**$_4$).[25] Mějme svět s a nějakou formulí φ, kterou nechceme, aby agent věděl. To znamená, že ji chceme stanovit jako neuvědomenou, tj. $\varphi \notin \mathcal{A}(s)$ pro svět s. Z toho vyplývá, že $(\mathcal{M}, s) \not\models A\varphi$, a tedy že $(\mathcal{M}, s) \not\models K^e\varphi$. Tímto způsobem lze snadno prokázat neplatnost (**LO**$_1$)–(**LO**$_4$). Princip tohoto řešení spočívá tedy v tom, že máme možnost libovolnou formuli označit jako uvědomělou či neuvědomělou. Všimněme si, že jde vlastně o syntaktickou analogii našeho řešení logické vševědoucnosti pomocí nemožných světů, neboť tam jsme naopak mohli označovat libovolné formule za pravdivé či nepravdivé.

Funkce $\mathcal{A}$ poskytuje značnou flexibilitu při vypořádávání se s logickou vševědoucností, nutno ovšem poznamenat, že chování operátoru A se někdy značně liší od našich intuic ohledně uvědomování, např. formule $A(\varphi \wedge \psi) \wedge \neg A(\psi \wedge \varphi)$ je splnitelná. To by se ale dalo ještě připustit u malých dětí, které mohou vidět rozdíl mezi $p \wedge q$ a $q \wedge p$. Problematičtější je ovšem to, že jsou splnitelné i takové formule jako např. $AA\varphi \wedge \neg A\varphi$. S další výtkou přišel Kurt Konolige.[26] Upozornil na to, že toto řešení vlastně omezuje inferenční schopnosti agenta až poté, co daná odvození provedl. Jinými slovy, agent je logicky vševědoucí do té doby, dokud není aplikována funkce $\mathcal{A}$. Pokud si to shrneme, v základu GAL leží syntaktické pojetí uvědomění, a v důsledku toho i znalosti. Logika uvědomění je tak velmi jemnozrnná: agent si může uvědomovat $\varphi \vee \psi$, aniž by si uvědomoval $\psi \vee \varphi$. Agent ovšem není schopen věřit nekonzistentním tvr-

[25] Srov. [Meyer & Hoek, 1995], s. 270.
[26] Srov. [Konolige, 1986].

zením a zůstává logicky vševědoucí v rámci explicitní znalosti.

5.2.5 Předsudky a slepá víra

Zajímavým způsobem přistoupili k řešení logické vševědoucnosti John-Jules Ch. Meyer a Wiebe van der Hoek.[27] Jejich řešení se po technické stránce velmi podobá GAL, ovšem místo funkce $\mathcal{A}$ sloužící k eliminaci či odfiltrování nechtěných znalostí zavádí tzv. *principy*, pomocí nichž můžeme v systému modelovat něco, co bychom mohli obecně nazvat předsudky či slepou vírou v určitá, třeba i nekonzistentní, přesvědčení. Hlavní myšlenka v pozadí je tedy ta, že člověk může posuzovat skutečnost, a tedy vytvářet si o ní určitá přesvědčení podle jistých pravidel neboli principů, kterých si třeba není ani sám vědom. Tím se vlastně dostáváme k novému druhu přesvědčení, které je ještě „implicitnější než implicitní".[28] Tuto formulaci je samozřejmě nutné brát s rezervou, nicméně přesně vystihuje to, o co tu van der Hoek a Meyer usiluje, tj. o zachycení nevědomých předsudků, které jsou uloženy hlouběji než ty implicitní. Tomuto novému druhu přesvědčení říkají van der Hoek a Meyer právě principy.[29] To, že má agent nějaké předsudky φ, zapíšeme jako $P\varphi$.

Nechť $\mathcal{L}^P$ je jazyk výrokové modální logiky obohacený o operátor principiálního přesvědčení K^P a $\mathcal{P}$ neprázdná množina atomických formulí, pak model s předsudky $\mathcal{M}$ má tvar $\langle S, \pi, \mathcal{R}, \mathcal{Q}\rangle$, kde:

[27] Srov. [van der Hoek, 1989].

[28] Van der Hoek a Meyer označují principy jako implicitní přesvědčení, nicméně jejich terminologii opouštíme, aby nedošlo k záměně s „běžným" implicitním přesvědčením (resp. znalostí), které jsme si představili v sekci 5.1.1.

[29] Všimněme si, že na tyto principy můžeme vlastně pohlížet jako na redukční principy, které jsme zmiňovali v sekci 4.1.4, ovšem s tím rozdílem, že jsou zohledněny samotným systémem SEL.

- S je neprázdná množina možných světů,
- π je funkce, která přiřazuje pravdivostní hodnotu atomickým formulím v možných světech, tj. $\pi : S \mapsto (\mathcal{P} \mapsto \{\texttt{true}, \texttt{false}\})$,
- $\mathcal{R}$ je relace dosažitelnosti,
- $\mathcal{Q}$ je funkce, která přiřazuje agentovi v každém světě $s \in S$ množinu formulí, které tvoří jeho principy, tj. $\mathcal{Q} : S \mapsto \wp(\mathcal{L}^P)$.

Sémantika zůstává nezměněna, jen je potřeba přidat definici pro nový operátor P, který bude označovat principy:

- $(\mathcal{M}, s) \models P\varphi \Leftrightarrow \varphi \in \mathcal{Q}(s)$.

Slovně: agent disponuje principem φ právě tehdy, když formule φ je prvkem množiny, kterou agentovi přiřazuje funkce $\mathcal{Q}$. Principiální přesvědčení je pak vymezeno pomocí principů:

- $(\mathcal{M}, s) \models K^P\varphi \Leftrightarrow \varphi \in \mathcal{Q}(s)$ nebo $(\mathcal{M}, t) \models \varphi$ pro všechna t taková, že $(s, t) \in \mathcal{R}$.

Agent je tedy o φ principiálně přesvědčen tehdy, když ví, že φ, nebo když je φ jeden z jeho principů, formálně $K^P\varphi \leftrightarrow (K\varphi \vee P\varphi)$.

Tento přístup umí řešit (**LO**$_1$), (**LO**$_2$), (**LO**$_3$), platnost si zachovává (**LO**$_4$).[30] Je zřejmé, že tento přístup je velmi blízký předcházejícímu řešení, a tudíž trpí i podobnými nedostatky. Na rozdíl od GAL ovšem dokáže zachytit i nekonzistentní přesvědčení, tj. formule tvaru $K^P(\varphi \wedge \neg\varphi)$ jsou splnitelné, pokud zvolíme např. $(\varphi \wedge \neg\varphi) \in \mathcal{Q}(s)$ pro nějaký svět s. Výhody tohoto přístupu lze tedy spatřovat v tom, že je schopen modelovat nekonzistentní přesvědčení, aniž by k tomu musel zavádět problematické entity jako nemožné či nekonzistentní a neúplné světy.

[30] Srov. [Meyer & Hoek, 1995].

5.2.6 Lokální uvažování

Fagin a Halpern nabídli ještě jedno řešení logické vševědoucnosti, které se snaží vyhnout jak nekonzistentním světům, tak i syntaktickému charakteru uvědomění, a to zavedením tzv. *svazků* (angl. clusters).[31] Hlavní myšlenka v pozadí je ta, že agent uvažuje v rámci různých vzájemně nezávislých svazků, přičemž svazek můžeme neformálně chápat jako určitý referenční rámec či kontext dané znalosti. Fagin a Halpern nazvali toto řešení jako *lokální uvažování* (angl. local reasoning). Agent tak může věřit jistému tvrzení v jednom svazku a současně jeho opaku v jiném svazku, přestože si je velmi dobře vědom toho, že obě tvrzení nemohou platit současně a že věří prakticky vzato v kontradikci. Svazky mu tedy umožňují mít vzájemně nekonzistentní názory. Jedná se v podstatně o takový orwellovský *doublethink*, tj. schopnost držet současně dva názory, které se vzájemně vylučují, vědět, o jejich protikladnosti, a přesto věřit v oba. Meyer a van der Hoek přichází s následujícím příkladem: fyzik může o elektronu uvažovat buď jako o částici, nebo jako o vlně v závislosti na tom, jestli o něm uvažuje v rámci fyziky klasické nebo kvantové.

Formálně je tento přístup vystaven na tzv. *svazkových* modelech, které byly inspirovány Montague-Scottovou sémantikou. Tu navrhli nezávisle na sobě Richard Montague a Dana Scott a jedná se o alternativní sémantiku možných světů.[32] Svazkové modely se od standardních Kripkeho modelů liší v tom, že namísto množin epistemických alternativ pracují s množinou množin epistemických alternativ. V rámci tohoto přístupu jsou tedy epistemické alternativy SEL rozděleny do podmnožin neboli svazků. Relace epistemické dosažitelnosti $\mathcal{R}$ je přitom zcela

[31] Srov. [Fagin & Halpern, 1988].

[32] Srov. [Montague, 1970]. Více o Montague-Scottově sémantice viz Dodatek A.

vynechána, neboť její roli v zásadě zastupuje právě množina svazků. To nám umožňuje považovat určitou formuli φ za znalost v jednom svazku, přičemž φ už nemusí být znalostí ve svazku jiném. Nutno dodat, že tyto svazky nemusí být vzájemně konzistentní.

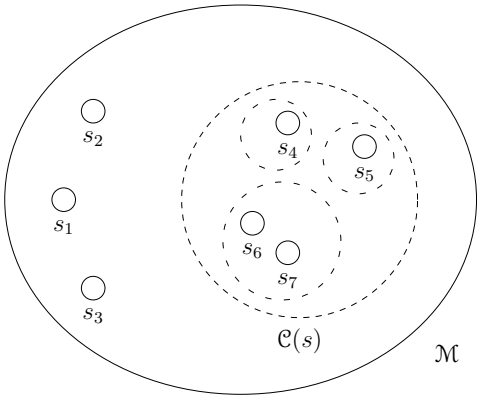
Pokud bychom se vrátili k našemu příkladu z fyziky, tyto svazky je možné chápat např. jako dvě teorie mechaniky, klasickou a kvantovou. Přestože jsou vzájemně nekonzistentní, je rozumné brát v úvahu obě. V logice lokálního uvažování je tedy agent nahlížen jako někdo, kde je schopen pracovat v rámci různých myšlenkových soustav (angl. frames of mind), přičemž každá soustava (svazek) je modelován pomocí odlišných epistemických alternativ. Nyní k formálnímu zachycení.

Nechť $\mathcal{L}$ je jazyk výrokové modální logiky a $\mathcal{P}$ neprázdná množina atomických formulí, pak svazkový model $\mathcal{M}$ má tvar $\langle S, \pi, \mathcal{C}\rangle$, kde:

- S je neprázdná množina možných světů,
- π je funkce, která přiřazuje pravdivostní hodnotu atomickým formulím v možných světech, tj. $\pi : S \mapsto (\mathcal{P} \mapsto \{\texttt{true}, \texttt{false}\})$,
- $\mathcal{C}$ je funkce $S \mapsto \wp(\wp(S))$ přiřazující agentovi množinu formulí, kterým v daném světě věří, přičemž $\mathcal{C}(s)$ je množina podmnožin S (viz obr. 5.3).

Nechť $c_1, ..., c_k$ jsou jednotlivé svazky, pak $\mathcal{C}(s) = \{c_1, ..., c_k\}$ znamená, že ve světě s agent věří svazkům $c_1, ..., c_k$. $\mathcal{C}(s)$ tedy indikuje ty svazky (tj. myšlenkové soustavy), které agent bere v potaz.

Sémantika pro atomické formule, negaci a konjunkci zůstane nezměněna, ovšem sémantiku operátoru K definujeme pomocí *pravdivostní množiny*. Nechť φ je libovolná formule, pak $\|\varphi\|^{\mathcal{M}}$ je pravdivostní množina formule φ v modelu $\mathcal{M}$, tj. množina světů, ve kterých je φ pravdivá.



Obrázek 5.3: Lokální uvažování

Pro libovolnou formuli φ tedy dostáváme, že její pravdivostní množina je $\|\varphi\|^{\mathcal{M}} = \{s \mid s \models \varphi\}$. Operátor K pak definujeme následovně:

- $(\mathcal{M}, s) \models K\varphi \Leftrightarrow \|\varphi\|^{\mathcal{M}} \in \mathcal{C}(s)$.

Slovně: agent ví, že φ, jestliže pravdivostní množina $\|\varphi\|^{\mathcal{M}}$ je prvkem $\mathcal{C}(s)$, tj. množiny podmnožin S ve světě s. O agentovi tak řekneme, že ví, že určitá určitá formule φ platí, jestliže je φ pravdivá ve všech světech, které agent považuje za možné v odpovídajícím svazku.

Tento koncept nám dále umožňuje definovat rozdíl mezi *lokální* (K^L) a *globální* znalostí (K^G).

- $(\mathcal{M}, s) \models K^L\varphi \Leftrightarrow$ je tu množina světů $T \in \mathcal{C}(s)$ taková, že pro všechna $t \in T$ platí, že $(\mathcal{M}, t) \models \varphi$,
- $(\mathcal{M}, s) \models K^G\varphi \Leftrightarrow$ ve všech množinách světů $T \in \mathcal{C}(s)$ platí, že $(\mathcal{M}, t) \models \varphi$.

Slovně: agent ví *lokálně*, že platí nějaké φ, jestliže φ je pravdivé ve všech možných světech v nějakém svazku T. A o agentovi řekneme, že ví φ *globálně*, jestliže je φ pravdivé ve všech alternativách všech svazků T, které považuje za možné.

Tento přístup umí řešit ($\mathbf{LO}_2$). Agent může vědět, že φ v rámci jednoho svazku a že $\varphi \to \psi$ v rámci jiného svazku, ale nikdy nemusí vzít v úvahu takový svazek, který obsahuje současně jak φ, tak $\varphi \to \psi$, tudíž formule jako $K\varphi \wedge K(\varphi \to \psi) \to K\psi$ nejsou platné. Princip ($\mathbf{LO}_3$) si avšak zachovává platnost (a pokud K nahradíme K^L, tak i ($\mathbf{LO}_1$) a ($\mathbf{LO}_4$)).[33] To je způsobeno právě tím, že agent je stále logicky vševědoucí v rozsahu konkrétního svazku, kde je jeho znalost i nadále uzavřena vzhledem k logické implikaci, ekvivalenci a tautologiím.

Shrnutí. Seznámili jsme se s šesti řešeními logické vševědoucnosti a pozornému čtenáři jistě neuniklo to, že ačkoliv se od sebe jednotlivá řešení na první pohled liší, podstata každého řešení zůstává stejná. Ať už šlo o nemožné světy (5.2.1), neúplné a nekonzistentní světy (5.2.2), nestandardní logiky (5.2.3), uvědomění (5.2.4), principy (5.2.5) nebo lokální uvažování (5.2.6), idea v pozadí byla vždy tatáž, a to zavedení takového prostoru, který bude mimo dosah zákonů klasické logiky. Jinými slovy, jednotlivá řešení se vždy opírala o zavedení nestandardních logik a upravovala samotné jádro systému SEL, což někdy vedlo až k opuštění celé Kripkeho sémantiky (jako např. v případě řešení 5.2.6).

Jednotlivá řešení se pak liší zejména mírou, s jakou modifikují výchozí systém SEL. Svým způsobem se tedy jedná o *ad hoc* řešení, která se s logickou vševědoucností vypořádávají na úkor jednoduchosti a ele-

[33] Srov. [Meyer & Hoek, 1995].

gantnosti výchozího systému SEL (Meyer např. mluví přímo o znečišťování[34]).

Dle Meyera a van der Hoeka tu ani žádné jiné řešení logické vševědoucnosti být nemůže:

> Nejobecnější metody řešení se zdají být příliš „nelogické“, kdežto ty specifičtější „logické“ metody si zase ponechávají slabší formy logické vševědoucnosti. Možná je čistě logické řešení problému logické vševědoucnosti *contradictio in terminis* a v konkrétních situacích se vždy musíme spokojit s výběrem toho modelu, který více či méně zachycuje takovou míru logické (ne)vševědoucnosti, která je v dané situaci přijatelná.[35]

K podobnému závěru nakonec dochází i Fagin et al.:

> Stejně jako si nemyslíme, že je tu jedna správná, pravá definice znalosti, která by zachycovala všechny nuance použití tohoto slova v angličtině, nemyslíme si ani, že je tu jen jeden sémantický přístup k vyřešení problému logické vševědoucnosti. (…) Volba řešení vždy nakonec závisí na konkrétní aplikaci.[36]

---

[34] Srov. [Meyer, 2001], s. 200.

[35] Srov. [Meyer & Hoek, 1995], s. 89. V originále: „The general methods seem to be too ‘illogical’, whereas the more special and ‘logical’ methods seem to suffer from some remnants of logical omniscience. Perhaps a perfectly logical solution to the logical omniscience problem is a *contradictio in terminis*, and we must satisfy ourselves in practical situations with a particular choice of the available models that more or less captures the extent of logical (non)omniscience that is accepttable in that particular situation.“ (překlad autor).

[36] Srov. [Fagin et al., 1995], s. 373–374. V originále: „Just as we do not feel that there is

Je třeba připomenout, že logická vševědoucnost nepředstavuje závažný problém, pokud SEL chápeme jako logiku modelující implicitní znalost nebo znalost idealizovaného agenta, resp. idealizovanou znalost. Nicméně obě tyto možnosti poněkud diskreditují epistemickou logiku v její původní snaze být epistemickou, tj. zabývat se znalostí nás lidí, neboť jak implicitní znalost, tak i idealizovaný agent jsou pouze teoretickými konstrukty. K problému logické vševědoucnosti bylo tedy třeba přistoupit jiným způsobem, konkrétně modifikací SEL.

one right, true definition of knowledge that captures all the nuances of the use of the word in English, we also do not feel, that there is a single semantic approach to deal with the logical-omniscience problem. (...) Ultimately, the choice of the approach used depends on the application." (překlad autor).

Kapitola 6

Epistemická logika dnes aneb kam dál?

Zde představená SEL (resp. SEL^{KB}, SEL^{n}, SEL^{nC}, SEL^{nCD} a SEL*) neboli kripkovská epistemologie, jak se jí někdy poněkud zavádějícím způsobem také přezdívá, je pouze jednou z možných epistemických logik. Je tu samozřejmě i celá řada alternativních přístupů opírající se např. o substrukturální logiky, vícehodnotové (fuzzy) logiky či hyperintenzionální logiky. Každý z těchto systémů přistupuje k řešení problému logické vševědoucnosti specifickým způsobem, který si na zbylém místě v krátkosti představíme.

6.1 Substrukturální logiky

Substrukturální logiky jsou neklasické logiky, které postrádají jedno nebo více tzv. *strukturálních pravidel*. Termín strukturální pravidla pochází z důkazově-teoretického (angl. proof-theoretic) zkoumání logiky a ozna-

čuje se jím sada podmínek, která charakterizuje klasickou relaci důsledku, tradičně zapisovanou pomocí symbolu $\vdash$. Skutečnost, že např. „z A je odvoditelné A“ pak zapíšeme jako $A \vdash A$. Mezi strukturální pravidla se řadí např. pravidlo oslabení (angl. weakening) **(PO)**, pravidlo kontrakce **(PK)** či pravidlo záměny **(PZ)**. Podrobněji:

$$\textbf{(PO)}\ \frac{\Gamma \vdash A}{\Gamma, B \vdash A}$$

Slovy: jestliže je z premis Γ odvoditelné A, pak z premis Γ a B musí být také odvoditelné A.[1]

$$\textbf{(PK)}\ \frac{\Gamma, B, B \vdash A}{\Gamma, B \vdash A}$$

Slovy: jestliže je z premis Γ, B, B odvoditelné A, pak musí být odvoditelné i z premis Γ, B. Jinými slovy, nezáleží na tom, kolikrát danou premisu použijeme.

$$\textbf{(PZ)}\ \frac{\Gamma, B, C \vdash A}{\Gamma, C, B \vdash A}$$

Pravidlo **(PZ)** nám říká, že nezáleží na pořadí premis.

Řešení logické vševědoucnosti, které si zde krátce představíme, navrhli Ondrej Majer and Michal Peliš[2] a opírá se o kripkovskou sémantiku pro substrukturální logiku. Hlavní myšlenkou v pozadí je nahrazení obecné znalostní modality K novou modalitou *potvrzení*, kterou budeme

[1] Čárka mezi Γ a B nalevo od $\vdash$ se běžně interpretuje jako notační zkratka za zápis $\Gamma \cup \{B\}$, tj. sjednocení původní množiny premis Γ s novou množinou obsahující B. Nutno ovšem dodat, že v rámci substrukturálních logik vstupují do hry i další interpretace, např. odmítnutí **(PK)** se chápe i jako zavedení předpokladu, že premisy tvoří tzv. multimnožinu (angl. multiset), tj. množinu, která může obsahovat opakující se prvky. Podobně, odmítnutí **(PZ)** se váže s předpokladem, že premisy tvoří posloupnost.

[2] Srov. [Majer & Peliš, 2013], dále také [Bílková et al., 2010]. Další přístup využívající substrukturální logiky je možné naleznout v [Sedlár, 2013] či [Sedlár, 2014].

značit ©. Místo možných světů zavedeme tzv. *informační stavy*, které mohou být neúplné i nekonzistentní. Z těch je dále vyčleněna neprázdná podmnožina L tzv. logických stavů, nad kterou je definováno částečné uspořádání $\leq$. Relaci epistemické dosažitelnosti $\mathcal{R}$ pak nahradí nová relace potvrzování $\mathcal{S}$ (resp. relace „být zdrojem“).

Nechť $\mathcal{L}$ je jazyk výrokové modální logiky a $\mathcal{P}$ neprázdná množina atomických formulí, pak substrukturální model $\mathcal{M}$ má tvar $\langle S, L, \pi, \leq, \mathcal{R}, \mathcal{K}, \mathcal{S}\rangle$, kde:

- S je neprázdná množina informačních stavů,
- $L \subseteq S$ je neprázdná podmnožina logických stavů,
- π je funkce, která přiřazuje pravdivostní hodnotu atomickým formulím v informačních stavech, tj. $\pi : S \mapsto (\mathcal{P} \mapsto \{\texttt{true}, \texttt{false}\})$, tak, že pokud $x < y$, pak $\pi(x, p) = 1$, jinak $\pi(y, p) = 1$,
- $\leq$ částečné uspořádání nad L,
- $\mathcal{R}$ je ternární relace relevance,
- $\mathcal{K}$ je binární relace kompatibility, přičemž
  - jestliže $\mathcal{R}(xyz)$, $x' \leq x$, $y' \leq y$ a $z \leq z'$, pak $\mathcal{R}(x'y'z')$,
  - $x \leq y$ právě tehdy, když existují logické stavy z a z' takové, že $\mathcal{R}(zxy)$ a $\mathcal{R}(xz'y)$,
  - $\mathcal{K}(x, y)$, $x' \leq x$ a $y' \leq y$, pak $\mathcal{K}(x', y')$,
  - jestliže $\mathcal{R}(xyz)$, pak ke každému z' ($\mathcal{K}(z'z)$) existuje y' ($\mathcal{K}(y'y)$) takové, že $\mathcal{R}(xz'y')$,
- $\mathcal{S}$ je binární relace potvrzování, přičemž $\mathcal{S}(s, x)$ právě tehdy, když $s < x$ a $\mathcal{K}(s, x)$.[3]

[3] Na relaci $\mathcal{S}$ mohou být kladeny i další podmínky, více viz [Bílková et al., 2010].

Sémantika nového operátoru je pak definována následovně:

- $(\mathcal{M}, s) \models ©\varphi$ právě tehdy, když existuje informační stav t takový, že $(\mathcal{M}, t) \models \varphi$ a $\mathcal{S}(t, s)$.

O agentovi tedy řekneme, že ví, že platí nějaké φ (resp. že ho má potvrzené), pokud je toto φ pravdivé alespoň v jednom z jeho dřívějších kompatibilních informačních stavů.

Tento přístup se umí vypořádat s ($\mathbf{LO}_2$) a ($\mathbf{LO}_4$), jelikož ani **(K)**, ani **(NEC)** v tomto epistemickém substrukturálním rámci neplatí. Platnost si ovšem zachovává vzhledem na slabší logiku ($\mathbf{LO}_1$) a ($\mathbf{LO}_3$).[4]

6.2 Fuzzy logiky

Obecně řečeno, fuzzy logiky[5] jsou vícehodnotové logiky, které pohlíží na pravdivostními hodnoty `true` a `false` (resp. `1` a `0`) jako na krajní body intervalu. To znamená, že fuzzy logika pracuje se *stupni* pravdivosti, které jí umožňují systematicky pracovat s takovými vágními termíny jako např. vysoký, starý atp. Hlavní myšlenku řešení, které navrhl Libor Běhounek a které si zde také představíme, vychází z analogie mezi tzv. *paradoxem hromady* a právě problémem logické vševědoucnosti.[6]

Paradox hromady nás staví před následující problém. Mějme nějakou formuli p_0, která je pravdivá (např. „Jedno zrnko písku hromadu netvoří“), dále mějme další formuli $p_n \to p_{n+1}$, která se rovněž zdá neproblematická (např. „Pokud k zrnku písku přidáme další zrnko, pak hromadu stále nemáme“). Z toho ovšem vyplývá, že formule p_n pro nějakou

[4] Srov. [Majer & Peliš, 2013].

[5] Srov. [Zadeh, 1965].

[6] Srov. [Běhounek, 2007].

velkou hodnotu n (např. $n = 10^9$) je rovněž pravdivá. Takový závěr je ale nepřijatelný, neboť miliarda zrnek písku už hromadu rozhodně tvořit bude.

Máme tu před sebou tedy přijatelné premisy, ale nepřijatelný závěr. Jak z toho ven? Obecně máme dva možné přístupy: buď popřít pravdivost některé z premis, nebo závěru.[7] Bohužel ani jedna z těchto možností není moc uspokojivá: jedno zrnko rozhodně není hromada, přidání jen jednoho zrnka navíc z „nehromady“ hromadu neudělá a miliarda zrnek rozhodně hromada je.

Fuzzy logika nabízí následující řešení. Druhou premisu (tj. $p_n \to p_{n+1}$) nebudeme považovat za pravdivou (tzn. nepřipíšeme jí pravdivostní hodnotu 1), ale pouze za pravdě velmi blízkou (což můžeme zachytit např. přidělením pravdivostní hodnoty 0.9999). Řešení pak spočívá v opakování úsudku pomocí induktivní premisy tzv. *silnou konjunkcí*, která s každou aplikací snižuje stupeň pravdivosti. To znamená, že s každým dalším odvozením (přidaným zrnkem písku) bude závěr méně a méně pravdivý (0.9998, 0.9997...), tj. blížit se k 0, resp. nepravdivosti.

Běhounek si všiml, že analogická situace probíhá vlastně i v případě logické vševědoucnosti, jen hromadící se kopu písku nahradíme znalostmi (resp. jejich odvozováním). Podstatou řešení je tedy to, že znalost začneme chápat jako *fuzzy modalitu*, s tím že např. explicitní znalosti budou mít pravdivostní hodnoty blízké 1, kdežto pravdivostní hodnoty znalostí vyžadujících náročné odvozování se naopak budou blížit k 0.

[7] Teoreticky je tu samozřejmě i třetí možnost, a to popřít korektnost odvozovacího pravidla zde užitého, ale vzhledem k tomu, že se jedná o modus ponens, tato možnost nepřichází moc v úvahu.

6.3 Hyperintenzionální logiky

Základní idea hyperintenzionální logiky spočívá ve vykročení za intenzionální systémy, podobně jako jsme dříve postoupili za systémy extenzionální.[8]

Pravděpodobně nejproblematičtější vlastností intenzionálních systémů, se kterou jsme se zde setkali, je pojetí významu jakožto propozice, tj. funkce z možných světů do pravdivostních hodnot, v důsledku čehož jsme dospěli k tomu závěru, že všechna pravdivá logická tvrzení mají stejný význam. To se ale při bližším prozkoumání zdá jako velmi diskutabilní předpoklad. Odmítneme-li toto pojetí významu, odmítne také intenzionální systémy. Jaký systém, resp. pojetí významu, ale vlastně hledáme? Vraťme se ještě jednou k úsudku (5”):

$$\text{(5'')}\ \frac{\begin{array}{c} K\ \big((p \wedge q) \supset p\big) \\ \big(\big((p \wedge q) \supset p\big) \leftrightarrow \big((q \wedge p) \supset p\big)\end{array}}{K\ \big((p \wedge q) \supset p\big)}$$

SEL nás nutí k tomuto z epistemického hlediska problematickému závěru proto, že formule $\big((p \wedge q) \supset p\big)$ a $\big((q \wedge p) \supset p\big)$ mají v rámci intenzionálních systémů stejný význam. Potíž tedy spočívá v tom, že pro intenzionální systémy jsou formule jako $\big((p \wedge q) \supset p\big)$ a $\big((q \wedge p) \supset p\big)$ sémanticky ekvivalentní, a tedy v propozičních postojích zaměnitelné *salva veritate*.

To ovšem znamená, že problém by měl být vyřešen, pokud nalezneme systém, který nám umožní rozlišit význam formule jako $\big((p \wedge q) \supset p\big)$ a $\big((q \wedge p) \supset p\big)$. Jinými slovy, hledáme takový systém, který je od sebe

[8] Více o hyperintenzionálních systémech může čtenář naleznout např. v [Cresswell, 1985], [Tichý, 1988], [Duží et al., 2010] či [Raclavský, 2009].

schopen rozlišit dvě formule (propozice, tvrzení), z nichž jsou obě pravdivé ve všech možných světech. Hledáme tedy takový systém, který poskytuje jemnější kriterium pro stejnost významu než je logická ekvivalence, tj. hledáme vlastně jemnější pojetí významu samotného. Jak vidíme, v závěru nám tedy nejde o nic jiného než o celkově jemnozrnnější analýzu významu tvrzení.

Jsou tu nějaké systémy, které by se o podobnou analýzu pokoušely? Odpověď zní ano a říká se jim hyperintenzionální. V hyperintezionálních systémech není význam tvrzení pojímán jako propozice (resp. množina možných světů, ve kterých je dané tvrzení pravdivé), ale jako strukturovaná entita, která odráží (ať už logickou nebo gramatickou) strukturu daného tvrzení.

Pokud bychom se např. pokusili úsudek (5") výše zachytit v rámci Tichého transparentní intenzionální logiky (TIL),[9] dostali bychom:

$$\text{(5'')}\ \frac{\begin{array}{c}\lambda w\,\lambda t\,\Big[\big[{}^0\text{Vědět}\ w\big]\ t\big]\ {}^0\text{Alenka}\ {}^0\big[{}^0\supset\ [{}^0\wedge\ p\ q]\ p\big]\Big] \\ \Big[{}^0\leftrightarrow\ \big[{}^0\supset\ [{}^0\wedge\ p\ q]\ p\big]\ \big[{}^0\supset\ [{}^0\wedge\ q\ p]\ p\big]\Big]\end{array}}{\cancel{\lambda w\,\lambda t\,\Big[\big[{}^0\text{Vědět}\ w\big]\ t\big]\ {}^0\text{Alenka}\ {}^0\big[{}^0\supset\ [{}^0\wedge\ q\ p]\ p\big]\Big]}}$$

což je neplatný úsudek, neboť ${}^0\big[{}^0\supset\ [{}^0\wedge\ p\ q]\ p\big]$ nese v TIL jinou sémantickou informaci než ${}^0\big[{}^0\supset\ [{}^0\wedge\ q\ p]\ p\big]$, a tudíž je nelze jednoduše zaměnit.

V důsledku tohoto pojetí významu umí hyperintenzionální systémy efektivně čelit např. ($\mathbf{LO}_3$), neboť formule $\big((p \wedge q) \supset p\big)$ a $\big((q \wedge p) \supset p\big)$ už nemají stejný význam. Nicméně samotná jemnozrnnost významu nám nijak nepomůže vypořádat s uzávěrem vzhledem k materiální implikaci —

[9] Srov. [Tichý, 1988], [Duží et al., 2010].

pokud známe premisy (ať už s jakkoli jemným významem), stále musíme znát i závěr, tudíž ($\mathbf{LO}_2$) neboli K-axiom zůstává stále v platnosti.[10]

6.4 Závěrečné poznámky

Cílem této knížky bylo zejména doprovodit čtenáře při jeho/jejích prvních krůčcích na poli epistemické logiky. Tento záměr se projevil zejména v záběru knihy, který je velmi limitovaný a omezený mnohdy jen na ty základní informace. To se vztahuje i na zde představená řešení logické vševědoucnosti, jež představují jen malou část z možných přístupů, jak se s ní vypořádat.

Jinak řečeno, SEL je jedním pokusem ze širší snahy formální epistemologie o vytvoření co možná nejadekvátnějšího modelu znalosti. A ačkoliv SEL rozhodně není dokonalá, je to stále velmi živý směr bádání, který právem přitahuje pozornost jak tradičních epistemologů (své má rozhodně co říci k tématům jako např. paradox poznatelnosti, Gettierův problém či falibilismus), tak i informatiků (zejm. témata spojená s multiagentními systémy).

Jedním z důvodů je nepochybně i to, že SEL nabízí velmi dobrý poměr „cena/výkon“ a značnou flexibilitu, co se různých modifikací týče (jak jsme mohli vidět např. v různých řešeních logické vševědoucnosti). Stručně řečeno, i přes množství nedostatků, problémů a komplikací má SEL nadále co nabídnout a je stále SEL v přísném slova smyslu, tj. standardní epistemickou logikou.

[10] O jeho odstranění se pokusili např. [Duží et al., 2004].

Dodatek A

Montague-Scottova sémantika

Montague-Scottovu sémantiku (známou angl. také jako *neighbourhood semantics*) můžeme chápat jako zobecnění Kripkeho sémantiky. Hlavním rozdílem je to, že v rámci této sémantiky není objektem znalosti propozice (resp. množina světů), ale množina propozic. Tomuto objektu se někdy také říká *okolí*[1] (angl. neighbourhood). O agentovi tedy řekneme, že zná nějakou formuli φ právě tehdy, když je φ pravdivá alespoň v jedné z těchto množin světů, které agent považuje za možné.

Nechť $\mathcal{L}$ je jazyk výrokové modální logiky a $\mathcal{P}$ neprázdná množina atomických formulí, pak model sousedství je trojice $\mathcal{M}$ tvaru $\langle S, \pi, \mathcal{V}\rangle$, kde:

- S je neprázdná množina možných světů,
- π je funkce, která přiřazuje pravdivostní hodnotu atomickým formulím v možných světech, tj. $\pi : S \mapsto (\mathcal{P} \mapsto \{\texttt{true}, \texttt{false}\})$,
- $\mathcal{V}$ je funkce $S \mapsto \wp(\wp(S))$, která každému světu přiřazuje množinu

[1] Srov. [Běhounek, 2005].

formulí, tj. těch formulí, kterým agent v daném světě $s \in S$ věří. $\mathcal{V}$ je tedy funkce z S do potenční množiny potenční množiny S.

Relace splňování $\models$ je definována podobně jako v Kripkeho sémantice, jediným rozdílem je definice znalostního operátoru:

- $(\mathcal{M}, s) \models K\varphi$ právě tehdy, když $\{t \in S \mid (\mathcal{M}, t) \models \varphi\} \in \mathcal{V}(s)$.

Slovy: agent ví, že φ platí ve světě s právě tehdy, když množina světů, ve kterých platí φ, je prvkem okolí světa s.

Tato sémantika tedy pro každý svět uvádí explicitně všechny formule, kterým agent v daném světě věří. Tím se okamžitě vyhneme jistým druhům logické vševědoucnosti, např. uzávěru vzhledem k logickému důsledku, neboť jednotlivé množiny světů nejsou uzavřeny vzhledem ke svojí nadmnožině, tj. agent může např. vědět φ v rámci jedné množiny možných světů, $\varphi \to \psi$ v rámci jiné množiny, ale nikdy nemusí vědět, že ψ, jelikož nemusí považovat za možnou takovou množinu možných světů, která obsahuje obě formule současně. Jinými slovy, tato sémantika netrpí ($\mathbf{LO}_2$). Na druhé straně, potíž stále představuje např. ($\mathbf{LO}_3$), která nemůže být v rámci této sémantiky oslabena.

Dodatek B

Syntaktická znalost

Syntaktický přístup k řešení logické vševědoucnosti se opírá o větné pojetí znalosti. To jsme sice ve třetí kapitole odmítli jako krajně problematické, ale přesto bude užitečné seznámit se s řešením logické vševědoucnosti, které je na tomto syntaktickém pojetí založeno. Hlavní myšlenka v pozadí je ta, že znalost agenta je určitý explicitní seznam formulí, které agent v daném světě zná. Pokud tam nějaká formule φ je, pak ji zná, a naopak. Agent tedy ví, že φ právě tehdy, když si pamatuje (má ve znalostní bázi) větu „vím, že φ“. Můžeme říci, že v rámci syntaktického přístupu je znalost chápána v podstatě jako větný predikát náležející právě těm větám, které jsou explicitně uvedeny v agentově rejstříku znalostí. Model znalosti se tak stává jen seznamem vět (tj. syntaktických útvarů), který není uzavřený vzhledem k žádnému logickému principu. Jednotlivé znalosti se tedy od sebe liší svým syntaktickým tvarem a φ a $\neg\neg\varphi$ může představovat dvě odlišné znalosti.[1]

Nechť $\mathcal{L}$ je jazyk výrokové modální logiky a $\mathcal{P}$ neprázdná množina

[1] Všimněme si, že syntaktickou znalost lze tedy pojímat jako fragment sémantiky „uvědomění“ představené v sekci 5.2.4.

atomických formulí, pak model syntaktické znalosti je model $\mathcal{M}$ tvaru $\langle S, \pi, \mathcal{Z} \rangle$, kde:

- S je neprázdná množina možných světů,
- π je funkce, která přiřazuje pravdivostní hodnotu atomickým formulím v možných světech, tj. $\pi : S \mapsto (\mathcal{P} \mapsto \{\texttt{true}, \texttt{false}\})$,
- $\mathcal{Z}$ je funkce, která světu přiřazuje množinu formulí $\mathcal{Z}(s)$.

Sémantika pro atomické formule, konjunkci a negaci je definována stejným způsobem jako v SEL (tj. jako ve standardním Kripkeho modelu) s jediným rozdílem, a tou je definice znalosti. V rámci syntaktického přístupu je znalost definována následovně:

- $(\mathcal{M}, s) \models K\varphi$ právě tehdy, když $\varphi \in \mathcal{Z}(s)$.

Slovně: agent ví, že φ právě tehdy, když φ je prvkem množiny formulí, kterou agentovi přiřazuje $\mathcal{Z}$. Funkce $\mathcal{Z}$ tedy formálně zachycuje naši ideu explicitního seznamu znalostí.

Je zřejmé, že za takovýchto podmínek nemusí být formule pod žádným uzávěrem a že se tak lze velmi snadno vyvarovat logické vševědoucnosti. Cena, kterou za to zaplatíme, je ovšem velmi vysoká. Zásadní slabinou tohoto přístupu je to, že podává jen popis báze znalostí daného agenta. Z tohoto hlediska se tedy jedná o velmi nezajímavou logiku. Nadále jsou samozřejmě platné i námitky, které jsme vůči syntaktickému pojetí znalosti vznesli ve třetí kapitole. Stručně řečeno, syntaktický přístup je příliš jemnozrnný, neboť ve své základní podobě považuje jakékoli dvě množiny vět, které nejsou skutečně znak od znaku identické, za odlišné sémantické entity, a tedy i za odlišné znalostní báze. Uvažme

disjunkci $\varphi \vee \psi$, v rámci syntaktické přístupu nemáme žádný důvod předpokládat, že $K(\varphi \vee \psi) \Leftrightarrow K(\psi \vee \varphi)$ bude platné, neboť $\varphi \vee \psi$ může být v jeho bázi znalostí, kdežto $\psi \vee \varphi$ nikoli.

Dodatek C

Quinův operátor znalosti

Řekli jsme si, že v rámci predikátové logiky dochází ke komplikacím, když se snažíme analyzovat frázi „vědět, že“ (resp. odpovídající propoziční postoj „x ví, že platí y“) jako predikát. Nabízí si následující otázka: pokud „vědět, že“ není predikát, resp. nelze jako predikát analyzovat, co to tedy je? Na tuto otázku se pokusil odpověď Willard V. O. Quine.

Quine[1] navrhuje abychom tvrzením jako „Alenka ví, že každý člověk je smrtelný.“ nepřipisovali obecnou formu $\mathrm{P}(a, b)$, kde P je binární predikát „vědět, že“, a Alenka a b tvrzení „každý člověk je smrtelný“, ale spíše formu $\mathrm{P}(a)$, kde a je Alenka a P je komplexní unární predikát (značený pomocí $\langle\,\rangle$) složený pomocí nového operátoru $\mathsf{Bel}[\quad]$.

S tímto novým operátorem můžeme tvrzení jako „Alenka ví, že každý člověk je smrtelný.“ analyzovat následovně:

(2a’) $\quad \langle\, \mathsf{Bel}\,[\forall x\big(\textsc{Člověk}(x) \rightarrow \textsc{Smrtelný}(x)\big)]\,\rangle\,(\textsc{Alenka})$

Stručně řečeno, základní myšlenka je ta, že operátor $\mathsf{Bel}[\quad]$ se přichytí

[1] Srov. [Quine, 1964], [Quine, 1986].

na konkrétní věty a vytvoří z nich unární predikáty, se kterými pak můžeme dále pracovat.

Pokusme se nyní analyzovat celý úsudek (2):

$$(2'') \quad \frac{\langle\ \mathsf{Bel}\ [\forall x\big(\text{Člověk}(x) \rightarrow \text{Smrtelný}(x)\big)]\ \rangle\ (\text{Alenka}) \qquad \langle\ \mathsf{Bel}\ [\text{Člověk}(\text{Sokrates})]\ \rangle\ (\text{Alenka})}{\cancel{\langle\ \mathsf{Bel}\ [\text{Smrtelný}(\text{Sokrates})]\ \rangle\ (\text{Alenka})}}$$

Tato analýza splňuje jak kriterium B, tj. zachovává všechny podstatné informace, tak i kriterium A, tj. správně rozpoznává, že se jedná o neplatný úsudek, avšak nelze ji obecně považovat za uspokojivou, neboť se spoléhá na problematické větné pojetí znalosti kritizované v sekci 3.3.

Summary

The book *An Introduction to Epistemic Logic* serves primarily as a introductory text to the study of epistemic logic specifically geared towards students of philosophy. Main attention is given to the so called Standard Epistemic Logic (SEL for short), which is a based on the epistemic logic presented in the pioneering book *Knowledge & Belief* (1962) by Jaakko Hintikka and its later developments.

In Chapter 1 we briefly discuss the logico-philosophical background of propositional attitudes starting with Frege's analysis of the so called belief-sentences („A believes p") and its connection to epistemic logic.

In Chapter 2 we present five exemplary epistemic inferences, which we use as case studies for analyses throughout the book, and determine their correctness from the epistemic point of view.

In Chapter 3 we introduce the concept of extensional systems, i.e., systems that respect the compositionality and the substitution principles. More specifically, we focus on the classical propositional and predicate logic and try to analyze epistemic inferences from the previous chapter in their respective frameworks.

In Chapter 4 we introduce the concept of intensional systems, i.e., systems where the compositionality and the substitution principles fail. Further on we define Standard Epistemic Logic (SEL) with Kripke-style semantics and introduce its various variants and extensions, such as e.g., SEL for multiple agents, SEL with common knowledge and distributive knowledge. Finally, we extend SEL to its first-order version, SEL*, and we try to analyze the five epistemic inferences from the Chapter 2 in this new framework.

In Chapter 5 the problem of logical omniscience will be generally discussed. Then we overview six most common solutions, namely solutions utilizing impossible possible worlds, explicit knowledge, nonstandard logic, awareness, principles and local reasoning, respectively.

In Chapter 6 we offer short summary and briefly discuss alternative approaches to the solution of logical omniscience (and generally epistemic logic), namely those put forward by substructural logics, fuzzy logics and hyperintensional logics.

Literatura

[Artemov & Kuznets, 2009] Artemov, S. & Kuznets, R. (2009). Logical omniscience as a computational complexity problem. In *Proceedings of the 12th Conference on Theoretical Aspects of Rationality and Knowledge*, TARK '09, s. 14–23, New York, NY, USA. ACM.

[Barcan, 1946] Barcan, R. C. (1946). A functional calculus of first order based on strict implication. *Journal of Symbolic Logic*, 11(1):1–16.

[Bílková et al., 2010] Bílková, M., Majer, O., Peliš, M., & Restall, G. (2010). Relevant agents. In *Advances in Modal Logic*, s. 22–38.

[Boh, 1993] Boh, I. (1993). *Epistemic Logic in the Later Middle Ages*. Topics in medieval philosophy. Routledge.

[Běhounek, 2005] Běhounek, L. (2005). Formální semantika logiky modalit. In Kolman, V., editor, *Možnost, skutečnost, nutnost: příspěvky k modální propedeutice*, s. 51–88. Filosofia.

[Běhounek, 2007] Běhounek, L. (2007). Dvě souvislosti mezi epistemickou a fuzzy logikou. In J. Kelemen, V. Kvasnička, J. P., editor, *Kognicia a umělé život VII*, s. 37–42. Slezská univerzita v Opavě.

[Carnap, 1947] Carnap, R. (1947). *Meaning and Necessity: A Study in Semantics and Modal Logic*. Number v. 3 in Meaning and Necessity: A Study in Semantics and Modal Logic. University of Chicago Press.

[Castañeda, 1964] Castañeda, H.-N. (1964). A note on S5. *Journal of Symbolic Logic*, 29(4):191–192.

[Chellas, 1980] Chellas, B. (1980). *Modal Logic: An Introduction*. Cambridge University Press.

[Chisholm, 1963] Chisholm, R. M. (1963). The logic of knowing. *Journal of Philosophy*, 60(25):773–795.

[Clark & Marshall, 1981] Clark, H. H. & Marshall, C. R. (1981). Definite reference and mutual knowledge. In Joshi, A., Weber, B. H., & Sag, I. A., editors, *Elements of Discourse Understanding*. Cambridge University Press.

[Cresswell, 1985] Cresswell, J. (1985). *Structured Meanings: The Semantics of Propositional Attitudes*. Bradford books. MIT Press.

[Dretske, 1970] Dretske, F. I. (1970). Epistemic operators. *Journal of Philosophy*, 67(24):1007–1023.

[Duc, 2001] Duc, H. N. (2001). *Resource-Bounded Reasoning about Knowledge*. PhD thesis, Faculty of Mathematics and Informatics, University of Leipzig.

[Duží et al., 2010] Duží, M., Jespersen, B., & Materna, P. (2010). *Procedural Semantics for Hyperintensional Logic: Foundations and Applications of Transparent Intensional Logic*. Logic, Epistemology, and the Unity of Science. Springer.

[Duží et al., 2004] Duží, M., Jespersen, B., & Müller, J. (2004). Epistemic closure and inferable knowledge. *The Logica Yearbook*, s. 125–140.

[Eberle, 1974] Eberle, R. (1974). A logic of believing, knowing, and inferring. *Synthese*, 26(3-4):356–382.

[Fagin & Halpern, 1988] Fagin, R. & Halpern, J. Y. (1988). Belief, awareness, and limited reasoning. *Artificial Intelligence*, 34:39–76.

[Fagin et al., 1995] Fagin, R., Moses, Y., Halpern, J., & Vardi, M. (1995). *Reasoning About Knowledge*. MIT Press.

[Frege, 1892] Frege, G. (1892). Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik*, 100:25–50.

[Frege, 1992] Frege, G. (1992). O smyslu a významu. *Scientia et Philosophia*, 4:33–75.

[Gochet & Gribomont, 2006] Gochet, P. & Gribomont, P. (2006). Epistemic logic. In Gabbay, D. & Woods, J., editors, *Handbook of the History of Logic. Vol 7. Logic and the Modalities in the Twentieth Century*, volume 7 of *Handbook of the History of Logic*, s. 101–195. Elsevier Science.

[Hales, 1995] Hales, S. D. (1995). Epistemic closure principles. *Southern Journal of Philosophy*, 33(2):185–202.

[Halpern & Moses, 1985] Halpern, J. Y. & Moses, Y. (1985). A guide to the modal logics of knowledge and belief: Preliminary draft. In *Proceedings of the 9th International Joint Conference on Artificial Intelligence - Volume 1*, IJCAI'85, s. 480–490, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

[Hendricks, 2006] Hendricks, V. (2006). *Mainstream and Formal Epistemology*. Cambridge University Press.

[Hintikka, 1962] Hintikka, J. (1962). *Knowledge and Belief: An Introduction to the Logic of the Two Notions*. Contemporary philosophy. Cornell University Press.

[Hintikka, 1975] Hintikka, J. (1975). Impossible possible worlds vindicated. *Journal of Philosophical Logic*, 4(4):475–484.

[Hintikka & Halonen, 1998] Hintikka, J. & Halonen, I. (1998). Epistemic logic. In Craig, E., editor, *Routledge Encyclopedia of Philosophy: A posteriori to Bradwardine. Vol. 1*. Routledge.

[Hintikka & Hintikka, 1989] Hintikka, J. & Hintikka, M. (1989). *The Logic of Epistemology and the Epistemology of Logic: Selected Essays*. Studies in epistemology, logic, methodology, and philosophy of science. Springer.

[Hocutt, 1972] Hocutt, M. O. (1972). Is epistemic logic possible? *Notre Dame Journal of Formal Logic*, 13(4):433–453.

[Holliday, 2013] Holliday, W. H. (2013). Epistemic logic and epistemology. *Handbook of formal philosophy. Dordrecht: Springer, forthcoming*.

[Kolman, 2005] Kolman, V. (2005). In Kolman, V., editor, *Možnost, skutečnost, nutnost: příspěvky k modální propedeutice*. Filosofia.

[Konolige, 1986] Konolige, K. (1986). What awareness isn't: A sentential view of implicit and explicit belief. In *Proceedings of the 1986 Conference on Theoretical Aspects of Reasoning About Knowledge*, TARK '86, s. 241–250, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

[Kraus & Lehmann, 1988] Kraus, S. & Lehmann, D. (1988). Knowledge, belief and time. *Theor. Comput. Sci.*, 58(1-3):155–174.

[Kripke, 1980] Kripke, S. (1980). *Naming and Necessity*. Library of philosophy and logic. Harvard University Press.

[Kripke, 1963] Kripke, S. A. (1963). Semantical considerations on modal logic. *Acta Philosophica Fennica*, 16(1963):83–94.

[Kripke, 1965] Kripke, S. A. (1965). Semantical analysis of modal logic II. Non-normal modal propositional calculi. In Addison, J. W., Tarski, A., & Henkin, L., editors, *The Theory of Models*. North Holland.

[Lemmon, 1957] Lemmon, E. J. (1957). New foundations for Lewis modal systems. *Journal of Symbolic Logic*, 22(2):176–186.

[Lemmon, 1967] Lemmon, E. J. (1967). If I know, do I know that I know? In Stroll, A., editor, *Epistemology: New Essays in the Theory Of Knowledge*. New York: Harper and Rowe.

[Lenzen, 1978] Lenzen, W. (1978). Recent work in epistemic logic. *Acta Philosophica Fennica*, 30:1–219.

[Lenzen, 1979] Lenzen, W. (1979). Epistemologische betrachtungen zu [S4, S5]. *Erkenntnis*, 14(1):33–56.

[Lenzen, 2004] Lenzen, W. (2004). Epistemic logic. In I. Niiniluoto, M. Sintonen, J. W., editor, *Handbook of Epistemology*, s. 963–983. Kluwer Academic Publishers.

[Levesque, 1984] Levesque, H. J. (1984). A logic of implicit and explicit belief. In *AAAI*, s. 198–202.

[Lewis & Langford, 1959] Lewis, C. & Langford, C. (1959). *Symbolic Logic*. Dover Book. Dover Publ.

[Lewis, 1969] Lewis, D. (1969). *Convention: a philosophical study*. Harvard University Press.

[Lewis, 1986] Lewis, D. (1986). *On the Plurality of Worlds*. B. Blackwell.

[Łoś, 1948] Łoś, J. (1948). Logiki wielowartościeowe a formalizacje funkcji intensyonalnych. *Kwartalnik filozoficzny*, s. 59–78.

[Majer & Peliš, 2013] Majer, O. & Peliš, M. (2013). Interpretace znalosti v substrukturálních rámcích. *Organon F*, Supplementary Issue 1:79–98.

[Makinson, 1965] Makinson, D. C. (1965). The paradox of the preface. *Analysis*, 25(6):205–207.

[Meyer, 2001] Meyer, J.-J. C. (2001). Epistemic logic. In Goble, L., editor, *The Blackwell Guide to Philosophical Logic*, s. 183–202. Oxford: Blackwell Publishers Ltd.

[Meyer & Hoek, 1995] Meyer, J.-J. C. & Hoek, W. V. D. (1995). *Epistemic Logic for AI and Computer Science*. Cambridge University Press, New York, NY, USA.

[Montague, 1970] Montague, R. (1970). Universal grammar. *Theoria*, 36(3):373–398.

[Nozick, 1981] Nozick, R. (1981). *Philosophical Explanations*. Harvard University Press.

[Peregrin, 2004] Peregrin, J. (2004). *Logika a logiky: systém klasické výrokové logiky, jeho rozšíření a alternativy*. Academia.

[Plantinga, 1974] Plantinga, A. (1974). *The Nature of Necessity*. Clarendon library of logic and philosophy. Clarendon Press.

[Priest, 2007] Priest, G. (2007). *Logika*. Průvodce pro každého. Dokořán.

[Prior, 1957] Prior, A. (1957). *Time and Modality*. Oxford University Press.

[Putnam, 1981] Putnam, H. (1981). *Reason, Truth and History*. Philosophical Papers. Cambridge University Press.

[Quine, 1943] Quine, W. V. (1943). Notes on existence and necessity. *Journal of Philosophy*, 40(5):113–127.

[Quine, 1964] Quine, W. V. O. (1964). *Word and Object*. MIT Press paperback series. MIT Press.

[Quine, 1986] Quine, W. V. O. (1986). *Philosophy of Logic*. Harvard University Press.

[Raclavský, 2009] Raclavský, J. (2009). *Jména a deskripce: logicko-sémantická zkoumání*. Olomouc: Nakladatelství, Olomouc.

[Rantala, 1982] Rantala, V. (1982). Impossible world semantics and logical omniscience. *Acta Philosophica Fennica*, 35:106–115.

[Rescher, 1960] Rescher, N. (1960). The problem of a logical theory of belief statements. *Philosophy of science*, s. 88–95.

[Rescher, 2006] Rescher, N. (2006). Epistemic logic. In Jacquette, D., editor, *A Companion to Philosophical Logic*, s. 478–490. Oxford: Blackwell Publishers Ltd.

[Sedlár, 2013] Sedlár, I. (2013). An outline of a substructural model of BTA belief. *Organon F*, Supplementary Issue 2:160–170.

[Sedlár, 2014] Sedlár, I. (2014). Epistemic extensions of modal distributive substructural logics. *(Vyjde v Journal of Logic and Computation)*.

[Sochor, 2011] Sochor, A. (2011). *Logika pro všechny ochotné myslet*. Univerzita Karlova.

[Stalnaker, 1984] Stalnaker, R. (1984). *Inquiry*. Bradford book. Bradford Book.

[Stalnaker, 1991] Stalnaker, R. (1991). The problem of logical omniscience, I. *Synthese*, 89(3):425–440.

[Stalnaker, 2006] Stalnaker, R. (2006). On logics of knowledge and belief. *Philosophical Studies*, 128(1):169–199.

[Švejdar, 2002] Švejdar, V. (2002). *Logika: neúplnost, složitost a nutnost*. Academia - nakladatelství Akademie věd ČR.

[Svoboda, 2010] Svoboda, V. (2010). *Logika a přirozený jazyk*. Filosofia.

[Thijsse, 1993] Thijsse, E. (1993). On total awareness logics. In de Rijke, M., editor, *Diamonds and Defaults*, volume 229 of *Synthese Library*, s. 309–347. Springer Netherlands.

[Tichý, 1988] Tichý, P. (1988). *The Foundations of Frege's Logic*. De Gruyter.

[Tversky & Kahneman, 1983] Tversky, A. & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, s. 293–315.

[van Benthem, 2006] van Benthem, J. (2006). Epistemic logic and epistemology: The state of their affairs. *Philosophical Studies*, 128(1):49–76.

[van der Hoek, 1989] van der Hoek, Meyer, J.-J. C. (1989). Possible logics for belief. *Logique et Analyse*, 32(127–128):177–194.

[van der Hoek, 1991] van der Hoek, W. (1991). Systems for knowledge and beliefs. In van Eijck, J., editor, *Logics in AI*, volume 478 of *Lecture Notes in Computer Science*, s. 267–281. Springer Berlin Heidelberg.

[van der Hoek, 1993] van der Hoek, W. (1993). Systems for knowledge and belief. *Journal of Logic and Computation*, 3(2):173–195.

[von Wright, 1951] von Wright, G. (1951). *An Essay in Modal Logic*. North-Holland publishing Company.

[Voorbraak, 1991] Voorbraak, F. (1991). The logic of objective knowledge and rational belief. In *Proceedings of the European Workshop on Logics in AI*, JELIA '90, s. 499–515, New York, NY, USA. Springer-Verlag New York, Inc.

[Voorbraak, 1992] Voorbraak, F. (1992). Generalized Kripke models for epistemic logic. In *Proceedings of the Fourth Conference on Theoretical Aspects of Reasoning About Knowledge*, TARK '92, s. 214–228, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

[Williamson, 2000] Williamson, T. (2000). *Knowledge and Its Limits*. Oxford Scholarship Online. Philosophy module. Oxford University Press.

[Zadeh, 1965] Zadeh, L. (1965). Fuzzy sets. *Information and Control*, 8:338–353.

Rejstřík

Epistemická logika:
úvod se zaměřením na studenty humanitních oborů

Ivo Pezlar

Vydala Masarykova univerzita v roce 2015
1. vydání

Grafický návrh obálky a sazba: Metoda spol. s r.o. (na základě podkladu autora), Hluboká 14, 639 00 Brno

Tisk: Tiskárna KNOPP, s. r. o., Kubelíkova 1224/42, 130 00 Praha 3

ISBN 978-80-210-7791-1

Tato kniha je určena každému, kdo se chce seznámit s epistemickou logikou (tj. logikou zabývající se formální analýzou pojmů *znalost* a *přesvědčení*) a problémy, které s sebou tato rychle se rozvíjející disciplína na pomezí filosofie, logiky a informatiky přináší. Hlavní pozornost bude věnována tzv. standardní epistemické logice (SEL), jejíž základy položil finský logik a filosof Jaakko Hintikka ve svém průkopnickém díle *Knowledge and Belief: An Introduction to the Logic of the Two Notions* (1962).

Mgr. Ivo Pezlar (nar. 12. 10. 1987 v Ivančicích) je doktorandem Katedry filosofie FF MU. Mezi hlavní oblasti jeho badatelského zájmu patří teorie důkazů, logická analýza přirozeného jazyka, epistemická logika a aplikace logiky v rámci umělé inteligence.

muni PRESS

ISBN 978-80-210-7791-1

9 788021 077911

esf evropský sociální fond v ČR
EVROPSKÁ UNIE
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY
OP Vzdělávání pro konkurenceschopnost
UNIVERSITAS MASARYKIANA BRUNENSIS
INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ