Neomylnosť a logika[1]

Igor Sedlár
Univerzita Komenského v Bratislave

Abstract: The paper presents preliminary results related to the logical analysis of the concept of infallibility. The concept is explicated using the resources of modal logic, namely in terms of the operator of possibility M and the operator of belief B. Infallibility with respect to *p* is explicated as impossibility of being wrong with respect to *p*. The infallibility operator is introduced: Iϕ is defined to mean the same as $\neg$M(Bϕ $\wedge$ $\neg\phi$). We prove various theorems about validity or invalidity of certain formulas with I in special classes of models for the combined language.

Keywords: infallibility, belief, possibility, modal logic.

Začnime jednou samozrejmosťou:

(1) Ľudia sa môžu mýliť.

Napriek tomu, že (1) znie skutočne triviálne, má v dejinách epistemológie dôležité miesto. Dejiny epistemológie môžeme totiž do určitej miery chápať ako rad pokusov ukázať, že (1) neplatí všeobecne. Inými slovami, ako rad pokusov dokázať nepravdivosť tvrdenia

(2) Ľudia sa môžu mýliť vo všetkom.

Jadrom rôznych odpovedí na skeptickú výzvu je totiž úsilie ukázať, že existujú propozície, pri ktorých sa *nemôžeme mýliť*. V Sedlár (2011) som načrtol spôsob formálneho skúmania takýchto propozícií a uviedol som niekoľko príkladov. V tomto príspevku sa pozriem na samotný pojem neomylnosti.

---

[1] Tento príspevok vznikol na Katedre logiky a metodológie vied FiF UK v Bratislave ako súčasť výskumného projektu *Sémantické modely, ich explanačná sila a aplikácie*, podporeného grantom VEGA č. 1/0046/11. Účastníkom sympózia ďakujem za diskusiu a užitočné pripomienky.

1 Neomylnosť: syntax

Je jasné, že *mýliť sa* vzhľadom na nejakú propozíciu p znamená byť presvedčený o pravdivosti p, keď p je v skutočnosti nepravdivá. Ak rozšírime jazyk klasickej výrokovej logiky o unárny operátor B, pričom Bp znamená „Som presvedčený, že p“, tak skutočnosť, že sa vzhľadom na p mýlim, vieme zachytiť formulou

(3) $\mathrm{B}p \wedge \neg p$

Jazyk s B však na formalizáciu pojmu neomylnosti nestačí. Byť neomylný vzhľadom na nejaké p znamená, že je *nemožné* mýliť sa vzhľadom na p. Ak rozšírime jazyk s operátorom B o nový unárny operátor M, pričom Mp znamená „Je možné, že p“, tak skutočnosť, že som neomylný vzhľadom na p, vieme zachytiť formulou

(4) $\neg\mathrm{M}\,(\mathrm{B}p \wedge \neg p)$

Ak v tomto rozšírenom jazyku známym spôsobom[2] definujeme operátor nevyhnutnosti L, tak formulu (4) vieme prepísať do ekvivalentnej a prehľadnejšej podoby

(5) $\mathrm{L}(\mathrm{B}p \rightarrow p)$

Kvôli prehľadnosti môžeme v rozšírenom jazyku s B a M definovať *operátor neomylnosti* I: Nech Ip je skratkou za (5) a podobne pre ľubovoľnú formulu ϕ na mieste p. Iϕ teda znamená „Som neomylný vzhľadom na ϕ“, kde ϕ je ľubovoľná formula.

2 Neomylnosť: sémantika

V tejto časti sa pozrieme na sémantiku jazyka s B a M . Tú postavíme na známych východiskách: na relačnej sémantike pre jazyk s M a na minimálnej sémantike pre jazyk s B. Časti 2.1 a 2.2 majú úvodný charakter a stručne vysvetľujú relačnú a minimálnu sémantiku výrokovej modálnej logiky a ich interpretácie, ktoré sú pre tento článok podstatné: metafyzickú a doxastickú. V časti 2.3 definujeme modely pre jazyk s B, M. Tie sú kombináciou relačných a minimálnych modelov.

[2] Teda Lϕ nech znamená $\neg\mathrm{M}\,\neg\phi$, pre ľubovoľnú formulu ϕ.

2.1 Možnosť

Modálna logika má v súčasnosti mnoho rôznych podôb,[3] no jej zárodkom boli systémy, medzi ktorých nedefinované pojmy patril aj pojem možnosti. Tento fakt môže byť vysvetlením skutočnosti, že modálna logika sa niekedy aj dnes pokladá primárne za „logiku možnosti a nevyhnutnosti". To zďaleka nie je adekvátne, avšak toto prihliadnutie na korene modálnej logiky môže pomôcť pri vysvetlení jej relačnej sémantiky.

Relačná sémantika výrokovej modálnej logiky vychádza z leibnizovského chápania možnosti ako pravdivosti v nejakom možnom svete a nevyhnutnosti ako pravdivosti v každom možnom svete. Kvôli širšej aplikovateľnosti táto sémantika využíva aj reláciu „dosiahnuteľnosti" medzi možnými svetmi, zvyčajne označovanú ako R. Východiskom tejto sémantiky je teda myšlienka, že výrok tvaru „Je možné, že *p*" je pravdivý v nejakom svete x práve vtedy, keď existuje nejaký svet y, ktorý je dosiahnuteľný z x a zároveň *p* je pravdivé v y. Doplňme, že výrok tvaru „Je nevyhnutné, že *p*" je pravdivý v x práve vtedy, keď *p* je pravdivé v každom y, ktorý je dosiahnuteľný z x.[4]

Vlastnosti R samozrejme vplývajú na platnosť výrokov. Napríklad výrok typu „Ak je nevyhnutné, že *p*, tak *p*" bude platný, ak predpokladáme, že R je reflexívna (teda pre každý svet x platí Rxx). Trocha voľnejšie môžeme povedať, že rôznym chápaniam nevyhnutnosti vyhovujú rôzne „nastavenia" relácie R. Pojem nevyhnutnosti sa však najčastejšie modeluje pomocou ekvivalenčnej relácie R (teda relácie, ktorá je reflexívna, symetrická a tranzitívna).

Sémantiku pre jazyk s M (ďalej „modálny jazyk") teraz definujeme presnejšie. *Modálny model* je trojica M = (W, R, V), kde W je neprázdna množina (množina možných svetov), R je reflexívna, symetrická a tranzitívna binárna relácia na W a V je funkcia, ktorá každej dvojici (ϕ, x), kde ϕ je formula modálneho jazyka a $x \in W$, priradí nejakú pravdivost-

---

[3] Pozri Blackburn – van Benthem – Wolter (2006).

[4] S modálnou logikou a jej sémantikou sa čitateľ môže bližšie zoznámiť napríklad v Peregrin (2004), Szomolányi (2003). Spomedzi zahraničných prác pozri napríklad Blackburn – de Rijke – Venema (2001), Hughes – Cresswell (1996), Chellas (1980) či van Benthem (2010).

nú hodnotu z množiny {0, 1}. V ďalej spĺňa nasledujúce požiadavky („vtt“ je skratka za „práve vtedy, keď“):

1. Pre každú výrokovú premennú p a každý svet $x \in W$ platí buď $V(p, x) = 1$, alebo $V(p, x) = 0$.
2. $V(\neg\phi, x) = 1$ vtt $V(\phi, x) = 0$
3. $V(\phi \wedge \psi, x) = 1$ vtt $V(\phi, x) = 1$ a $V(\psi, x) = 1$
4. $V(M\phi, x) = 1$ vtt $V(\phi, y) = 1$, pre nejaké také $y \in W$, že Rxy

Formula ϕ je *platná* v modálnom modeli $M = (W, R, V)$ práve vtedy, keď $V(\phi, x) = 1$ pre všetky $x \in W$. Keď definujeme $L\phi$ ako $\neg M \neg\phi$, tak $V(L\phi, x) = 1$ vtt $V(\phi, y) = 1$ pre všetky také $y \in W$, že Rxy. *Mod* nech je trieda všetkých modálnych modelov. Log(*Mod*) nech je trieda všetkých formúl modálneho jazyka platných v každom $M \in$ *Mod*.

Triedu Log(*Mod*) môžeme chápať ako súbor tvrdení, ktoré v istom zmysle popisujú pojem možnosti (a teda aj pojem nevyhnutnosti). Bude užitočné uviesť niekoľko formúl, ktoré do Log(*Mod*) patria:

(6) $L(p \rightarrow q) \rightarrow (Lp \rightarrow Lq)$
(7) $Lp \rightarrow p$
(8) $Lp \rightarrow LLp$
(9) $p \rightarrow MLp$
(10) $Mp \rightarrow LMp$

Spomeňme, že každá spomedzi týchto formúl je platná v každej triede kombinovaných modelov, o ktorých budeme hovoriť nižšie. Rovnako je platná aj každá substitučná inštancia týchto formúl. Platnosť niektorých z nich bude dôležitým predpokladom v dôkazoch uvedených nižšie.

2.2 Logika presvedčenia

Slovné spojenia „Je možné, že …“ a „Som presvedčený, že …“ majú zaujímavé spoločné črty. Po prvé, oba môžeme chápať tak, že označujú výrokové operátory. Napríklad, ak za … v „Som presvedčený, že …“ dosadím nejaký výrok, tak dostanem znova nejaký výrok. Po druhé, oba spomenuté výrokové operátory sú neextenzionálne. Pravdivostná hodnota výrokov typu „Je možné, že *p*“ a „Som presvedčený, že *p*“ nie je určená výlučne pravdivostnou hodnotou výroku *p*.

Ihneď po formulovaní relačnej sémantiky pre modálnu logiku bolo zrejmé, že túto sémantiku je možné interpretovať doxasticky, a teda ju použiť ako formálny model pojmu presvedčenia; pozri klasickú prácu

Hintikka (1962). Takto vniknuté formálne modely pojmu presvedčenia (a taktiež pojmu poznania) našli aplikácie v rôznych oblastiach.[5]

Z rôznych dôvodov[6] sa však doxasticky interpretovaná relačná sémantika zdá byť príliš idealizujúcim modelom presvedčení. Ak opustíme niektoré diskutabilné predpoklady, dostávame sa od relačnej sémantiky k minimálnej sémantike.[7]

Minimálna sémantika pre jazyk s operátorom B (ďalej „doxastický jazyk") je nasledovná. *Minimálny doxastický model* (ďalej iba „doxastický model") je trojica M = (W, N, V), kde W a V sú ako v relačnej sémantike, pričom N je funkcia, ktorá každému $x \in W$ priradí nejakú množinu podmnožín W. N(x) je teda množina množín prvkov W. (Nemusí však ísť o množinu všetkých množín prvkov W!) V ďalej spĺňa nasledujúce požiadavky:

1. Pre každú výrokovú premennú *p* a každý svet $x \in W$ platí buď $V(p, x) = 1$, alebo $V(p, x) = 0$
2. $V(\neg\phi, x) = 1$ vtt $V(\phi, x) = 0$
3. $V(\phi \wedge \psi, x) = 1$ vtt $V(\phi, x) = 1$ a $V(\psi, x) = 1$
4. $V(B\phi, x) = 1$ vtt $|\phi| \in N(x)$, kde $|\phi| = \{y : V(\phi, y) = 1\}$

Formula ϕ je *platná* v doxastickom modeli M = (W, R, V) práve vtedy, keď $V(\phi, x) = 1$ pre všetky $x \in W$.

Prečo by sme však mali súhlasiť s tým, že minimálna sémantika reprezentuje pojem presvedčenia? Po prvé, musíme si uvedomiť, že ak prvky W chápeme ako možné svety, tak množiny prvkov W reprezentujú propozície. Funkciu N potom môžeme chápať tak, že pre nejakého agenta stanovuje, o akých propozíciách je presvedčený v rôznych možných svetoch. N(x) teda reprezentuje množinu propozícií, ktorým agent verí vo svete x.

Minimálna sémantika so sebou nesie iba jeden predpoklad. Ak totiž formulám ϕ a ψ zodpovedá v každom modeli tá istá propozícia (teda ak sú logicky ekvivalentné), tak v každom modeli platí aj to, že agent je presvedčený o ϕ práve vtedy, keď je presvedčený o ψ. Minimálna

---

5 K modálno-logickému modelovaniu pojmov presvedčenia a poznania pozri Fagin – Halpern – Moses – Vardi (1995), Meyer – van der Hoek (1995), kapitoly 12 a 13 práce van Benthem (2010) či prehľadový článok Meyer (2001).

6 Pozri najmä 9. kapitolu práce Fagin – Halpern – Moses – Vardi (1995).

7 K minimálnej sémantike pozri kapitoly 7 – 9 práce Chellas (1980).

sémantika je teda vhodným nástrojom modelovania presvedčení agentov, ktorí sú schopní zistiť, že dve propozície sú logicky ekvivalentné a prispôsobiť tomu svoje presvedčenia.[8] Ide skutočne o minimálnu požiadavku. Je totiž jasné, že sám projekt formálneho modelovania presvedčení vyžaduje, aby sa presvedčenia riadili aspoň nejakými všeobecnými pravidlami. (Ak by sa žiadnymi neriadili, neexistovali by žiadne pravidlá, ktoré by formálne modelovanie mohlo odhaliť.)

Presvedčeniam však môžeme pripisovať rôzne dodatočné vlastnosti. Nasledujúca tabuľka uvádza niektoré vlastnosti, ktoré sa v literatúre často objavujú. Pri každej vlastnosti uvádzame formulu doxastického jazyka, ktorá ju reprezentuje a zároveň vysvetľuje.

Vlastnosť presvedčenia	**Formula**
Konzistentnosť	$\neg(B\phi \wedge B\neg\phi)$
Uzavretosť na konjunkcie	$(B\phi \wedge B\psi) \rightarrow B(\phi \wedge \psi)$
Distribúcia do konjunkcie	$B(\phi \wedge \psi) \rightarrow (B\phi \wedge B\psi)$
Pozitívna introspekcia	$B\phi \rightarrow BB\phi$
Negatívna introspekcia	$\neg B\phi \rightarrow B\neg B\phi$
Logicky dokonalé presvedčenie	$B(p \vee \neg p)$

Možné dodatočné predpoklady týkajúce sa presvedčenia vieme v minimálnej sémantike zohľadniť špecifickými „nastaveniami“ funkcie N. V nasledujúcej tabuľke uvádzame formuly z predchádzajúcej tabuľky, pričom pri každej uvádzame vlastnosť N, ktorá podmieňuje platnosť danej formuly ($A, B \subseteq W$).

Formula	**Vlastnosť N**
$\neg(B\phi \wedge B\neg\phi)$	Ak $A \in N(x)$, tak $(W - A) \notin N(x)$
$(B\phi \wedge B\psi) \rightarrow B(\phi \wedge \psi)$	Ak $A, B \in N(x)$, tak $A \cap B \in N(x)$
$B(\phi \wedge \psi) \rightarrow (B\phi \wedge B\psi)$	Ak $A \in N(x)$ a $A \subseteq B$, tak $B \in N(x)$
$B\phi \rightarrow BB\phi$	Ak $A \in N(x)$, tak $\{y : A \in N(y)\} \in N(x)$
$\neg B\phi \rightarrow B\neg B\phi$	Ak $A \notin N(x)$, tak $\{y : A \notin N(y)\} \in N(x)$
$B(p \vee \neg p)$	$W \in N(x)$

[8] Všimnime si, že minimálna sémantika nevyžaduje, aby bol agent schopný odhaliť všetky logické dôsledky svojich presvedčení. Požaduje sa iba „citlivosť na logickú ekvivalentnosť“.

2.3 Logika možnosti a presvedčenia

Sémantika pre jazyk, ktorý obsahuje M spolu s B (ďalej „dm-jazyk“) vznikne zlúčením relačných modelov (pre M) s minimálnymi modelmi (pre B). Pri minimálnych modeloch sme uvažovali o rôznych možných vlastnostiach N, takže aj pri kombinovaných modeloch bude nutné odlíšiť viacero druhov.

Definícia 2.1. *Základný* dm-model (skrátene „základný model“) je M = (W, R, N, V), kde W je neprázdna množina (množina možných svetov), R je reflexívna, symetrická a tranzitívna binárna relácia na W, N je funkcia z W do P (P(W)) (do množiny všetkých podmnožín W) a V je funkcia, ktorá každej dvojici (ϕ, x), kde ϕ je formula dm-jazyka a x ∈ W, priradí nejakú pravdivostnú hodnotu z množiny {0, 1}. V ďalej spĺňa nasledujúce požiadavky:

1. Pre každú výrokovú premennú p a každý svet x ∈ W platí buď V(p, x) = 1, alebo V(p, x) = 0
2. V(¬ϕ, x) = 1 vtt V(ϕ, x) = 0
3. V(ϕ ∧ ψ, x) = 1 vtt V(ϕ, x) = 1 a V(ψ, x) = 1
4. V(M ϕ, x) = 1 vtt V(ϕ, y) = 1 pre nejaké také y ∈ W, že Rxy
5. V(Bϕ, x) = 1 vtt |ϕ| ∈ N (x), kde tradične |ϕ| = {y : V(ϕ, y) = 1}

Formula ϕ je platná v základnom modeli M = (W, R, N, V) práve vtedy, keď V(ϕ, x) = 1 pre všetky x ∈ W. $MBel_E$ nech je množina základných modelov pre dm-jazyk. Log($MBel_E$) nech je trieda formúl dm-jazyka, ktoré sú platné v každom M ∈ $MBel_E$.

Definícia 2.2. *Racionálny* dm-model je taký základný model, v ktorom pre každú A ⊆ W navyše platí: Ak A ∈ N(x), tak (W – A) ∉ N(x). Triedu všetkých racionálnych modelov označíme $MBel_D$ a triedu všetkých formúl dm-jazyka platných v každom M ∈ $MBel_D$ označíme Log($MBel_D$).

Konjunktívny dm-model je taký základný model, v ktorom pre každú A ⊆ W navyše platí: Ak A, B ∈ N(x), tak A ∩ B ∈ N(x). Triedu všetkých konjunktívnych modelov označíme $MBel_C$ a triedu všetkých formúl dm-jazyka platných v každom M ∈ $MBel_C$ označíme Log($MBel_C$).

Regulárny dm-model je taký konjunktívny model, v ktorom pre každé A, B ⊆ W navyše platí: Ak A ∈ N(x) a A ⊆ B, tak B ∈ N(x). Triedu všetkých regulárnych modelov označíme $MBel_R$ a triedu všetkých formúl dm-jazyka platných v každom M ∈ $MBel_R$ označíme Log($MBel_R$).

Extrémny dm-model je taký regulárny model, v ktorom pre každú $A \subseteq W$ navyše platí:

1. Ak $A \in N(x)$, tak $(W - A) \notin N(x)$
2. Ak $ran_R(x) \subseteq A$, tak $A \in N(x)$, kde $ran_R(x) = \{y : Rxy\}$

Triedu všetkých extrémnych modelov označíme $MBel_X$ a triedu všetkých formúl dm-jazyka platných v každom $M \in MBel_X$ označíme $Log(MBel_X)$.

3 Vlastnosti neomylnosti

V tejto časti dokážeme niekoľko tvrdení o platnosti respektíve neplatnosti rôznych formúl v rôznych druhoch dm-modelov. Pripomeňme, že $I\phi$ je skratkou za $L(B\phi \rightarrow \phi)$, teda za $\neg M(B\phi \wedge \neg\phi)$.

3.1 Základné modely

Základné modely reprezentujú kombináciu pojmu možnosti s minimálnym pojmom presvedčenia. Formuly obsahujúce operátor neomylnosti, ktoré patria do $MBel_E$ teda v istom zmysle predstavujú „základné zákony“ neomylnosti.

Tvrdenie 3.1. Nasledujúce formuly patria do $Log(MBel_E)$:

(a) $(Bp \wedge Ip) \rightarrow p$
(b) $Ip \rightarrow LIp$
(c) $MIp \rightarrow Ip$
(d) $(MIp \wedge Bp) \rightarrow p$

Bod a) môžeme chápať ako prvý test vhodnosti modálnej reprezentácie neomylnosti. Ide totiž o tvrdenie, ktoré by o neomylnosti určite malo platiť: Ak som presvedčený, že p je pravdivé a zároveň som neomylný vzhľadom na p, tak p.

Bod b) sa na prvý pohľad môže zdať trocha prekvapujúci: Ak som neomylný vzhľadom na p, tak som nevyhnutne neomylný vzhľadom na p. Podľa tohto tvrdenia neomylnosť nie je kontingentná. Ak som neomylný vzhľadom na nejakú propozíciu, tak nie je („metafyzicky“) možné, aby som taký nebol. Po zoznámení sa s dôkazom tohto tvrdenia (pozri dodatok) však bude jasné, že platnosť uvedenej formuly je triviálnym dôsledkom skutočnosti, že neomylnosť je v podstate modálna vlastnosť: Ip znamená $L(Bp \rightarrow p)$. Pre nevyhnutnosť však platí $Lq \rightarrow LLq$.

Podobne je možné komentovať aj bod c). Ten sa na prvý pohľad tiež môže zdať trocha prekvapivý: Ak je možné, že som neomylný vzhľadom na p, tak som skutočne neomylný vzhľadom na p. Tu si však treba uvedomiť, podobne ako pri bode b), že platnosť uvedenej formuly je triviálnym dôsledkom predpokladaných čŕt „metafyzickej" možnosti. Stačí si uvedomiť, že každá formula tvaru $\mathrm{ML}\phi \rightarrow \mathrm{L}\phi$ je platná.

Podľa bodu d) platí, že ak je možné, že som neomylný vzhľadom na nejaké p, o ktorom som zároveň presvedčený, tak p je pravdivé. Toto silne znejúce tvrdenie je jednoduchým dôsledkom bodov a) a c) (pozri dodatok).

Prv, než pôjdeme ďalej, je dobré uvedomiť si, že neomylnosť neimplikuje pravdivosť:

Tvrdenie 3.2. Formula $\mathrm{I}p \rightarrow p$ nepatrí do Log($MBel_{\mathrm{E}}$).

Uvedená formula nie je tiež platná v žiadnej spomedzi tried modelov, ktorým sa venujeme nižšie.

3.2 Racionálne modely

Podstatnou črtou racionálnych modelov je to, že ak $|\phi| \in N(x)$, tak $|\neg\phi| \notin N(x)$. Inými slovami, ak je niekto presvedčený o ϕ, tak nie je presvedčený o $\neg\phi$. Tieto modely teda reprezentujú iba *konzistentné* presvedčenia. Často sa predpokladá, že konzistentnosť presvedčení je nevyhnutnou podmienkou ich racionálnosti[9] (odtiaľ názov tejto triedy modelov).

Je však zrejmé, že množina propozícií, o ktorých je niekto neomylný, nemusí byť sama konzistentná:

Tvrdenie 3.3. Formula $\mathrm{I}p \rightarrow \neg\mathrm{I}\neg p$ nepatrí do Log($MBel_{\mathrm{D}}$).

Platnosť tohto tvrdenia by čitateľa nemala prekvapiť. Predpoklad, že niekto je neomylný vzhľadom na „Bolí ma zub", nevylučuje možnosť, že by bol neomylný aj vzhľadom na „Nebolí ma zub". Často sa dokonca predpokladá, že napríklad z neomylnosti „Bolí ma zub" vyplýva neomylnosť vzhľadom na negáciu tohto výroku, teda vzhľadom na „Nebolí ma zub". Dá sa však ukázať, že tak to vo všeobecnosti nie je:

Tvrdenie 3.4. Formula $\mathrm{I}p \rightarrow \mathrm{I}\neg p$ nepatrí do Log($MBel_{\mathrm{D}}$).

---

9 Avšak pozri Makinson (1965).

Toto tvrdenie si zaslúži určitú pozornosť. Ak sa vo filozofickej literatúre stretneme s argumentmi v prospech názoru, že pri tvrdeniach určitého druhu sa nemôžeme mýliť, pričom daný argument pracuje výlučne s pozitívnymi tvrdeniami daného druhu,[10] tak daný argument nie je platný. Stále sa môžeme mýliť pri negatívnych tvrdeniach daného druhu.[11]

Je tiež zaujímavé si všimnúť, že formula $\mathrm{I}p \rightarrow \mathrm{I}\neg p$ nie je platná v žiadnej spomedzi tried modelov, o ktorých hovoríme nižšie.

3.3 Konjunktívne modely

Špeciálnou črtou konjunktívnych modelov je to, že v nich platí každá formula formy $(\mathrm{B}\phi \wedge \mathrm{B}\psi) \rightarrow \mathrm{B}(\phi \wedge \psi)$. Tieto modely teda reprezentujú predpoklad, že presvedčenie je uzavreté na konjunkcie: Ak som presvedčený o ϕ, ψ, tak som presvedčený aj o ich konjunkcii, teda $\phi \wedge \psi$.

Zdá sa rozumné očakávať, že táto črta presvedčení sa prenesie aj na neomylnosť. Tak to však nie je:

Tvrdenie 3.5. Formula $(\mathrm{I}p \wedge \mathrm{I}q) \rightarrow \mathrm{I}(p \wedge q)$ nepatrí do $\mathrm{Log}(MBel_C)$.

Toto tvrdenie ukazuje, že „uzavretosť na konjunkcie" sa z presvedčenia neprenáša na neomylnosť. To, že som neomylný vzhľadom na ϕ, ψ, ešte neznamená, že som neomylný vzhľadom na $\phi \wedge \psi$.

3.4 Regulárne modely

Význačnou črtou regulárnych modelov je to, že v nich pre každé ϕ, ψ platí $(\mathrm{B}\phi \wedge \mathrm{B}\psi) \leftrightarrow \mathrm{B}(\phi \wedge \psi)$. Reprezentujú teda predpoklad, že o konjunkcii dvoch tvrdení som presvedčený práve vtedy, keď som presvedčený o oboch daných tvrdeniach.

---

10 Tvrdeniami, ktoré nie sú ekvivalentné so žiadnym tvrdením, ktoré začína nepárnym počtom negácií.

11 Zaujímavým príkladom je propozícia vyjadrená výrokom „Existuje propozícia, vzhľadom na ktorú sa mýlim". Čitateľ si môže jednoducho overiť, že pri tejto propozícii sa nemôžem mýliť. Určite sa však môžem mýliť pri jej negácii, teda pri propozícii vyjadrenej výrokom „Nemýlim sa pri žiadnej propozícii". Pozri aj Sedlár (2011). Ešte jednoduchším, ba až triviálnym príkladom sú tautológie klasickej výrokovej logiky. Pri žiadnej z nich sa nemôžem mýliť, no určite sa môžem mýliť pri ich negáciách.

Dalo by sa čakať, že sa táto črta prenesie na neomylnosť (podobne ako to bolo s uzavretosťou na konjunkcie pri konjunktívnych modeloch). Tak to však nie je.

Tvrdenie 3.6. Nasledujúce formuly nepatria do Log($MBel_R$):
(a) $I(p \wedge q) \to (Ip \wedge Iq)$
(b) $I(p \vee q) \to (Ip \vee Iq)$
(c) $(Ip \vee Iq) \to I(p \vee q)$

Dá sa však ukázať, že platnosť formuly $B(p \wedge q) \to (Bp \wedge Bq)$ zaručuje skutočnosť, že neomylnosť je uzavretá na konjunkcie:

Tvrdenie 3.7. Formula $(Ip \wedge Iq) \to I(p \wedge q)$ patrí do Log($MBel_R$).

Zaujímavé je, že na platnosť tejto formuly nestačí predpoklad platnosti $(Bp \wedge Bq) \to B(p \wedge q)$. Musíme predpokladať opačnú implikáciu, teda $B(p \wedge q) \to (Bp \wedge Bq)$.

3.5 Extrémne modely

Ako už napovedá ich názov, extrémne modely pracujú s určitými veľmi silnými predpokladmi. Všimnime si, že pracujú s predpokladom, že modelované presvedčenia sú konzistentné. Ďalej pracujú so silnejším predpokladom, podľa ktorého platí, že ak $V(x, L\phi) = 1$, tak $|\phi| \in N(x)$. Inými slovami, ak je ϕ nevyhnutne pravdivá, tak je o nej agent presvedčený. Ešte jednoduchšie, agent je presvedčený o každej nevyhnutnej pravde.

Extrémnosť takýchto modelov však môžeme „zmierniť" tým, že oslabíme interpretáciu operátora B. Predpokladajme, že pri práci s extrémnymi modelmi nemodelujeme „uvedomelé" presvedčenia daného agenta, ale množiny tvrdení, ktoré z jeho „uvedomelých" presvedčení vyplývajú.[12] Pri takomto čítaní nie je uvedená črta extrémnych modelov taká extrémna: Z presvedčení daného agenta samozrejme vyplýva každá nevyhnutná pravda. (Práve preto, že je nevyhnutná. Inými slovami, nemôže byť nepravdivá.)

Predpokladajme teda, že modelujeme agentov s konzistentnými presvedčeniami, pričom $B\phi$ chápeme ako „Z presvedčení daného agenta vyplýva ϕ". Je jasné, že pri takomto čítaní sa zmení aj čítanie operátora I. Pri novom čítaní $I\phi$ znamená, že nie je možné, aby z presvedčení

---

[12] V literatúre sa v tejto súvislosti hovorí o „implicitných" presvedčeniach.

daného agenta vyplývalo ϕ v prípade, že ϕ je nepravdivé. Inými slovami, Iϕ je pravdivé v x, ak v ľubovoľnom „dosiahnuteľnom“ svete y má daný agent také presvedčenia, z ktorých vyplýva ϕ iba v prípade, ak je ϕ v danom y skutočne pravdivé. Ide o slabšie chápanie neomylnosti, no iste ide o určitú formu neomylnosti. Označme ju neomylnosť*.

Tvrdenie 3.8. Nasledujúce formuly patria do Log($MBel_X$):
(a) $\mathrm{BI}p \rightarrow \mathrm{I}p$
(b) $(\mathrm{BI}p \wedge \mathrm{B}p) \rightarrow p$

Ide o prekvapivo silné tvrdenie. Podľa bodu a) sa nemôže stať, že by z mojich konzistentných presvedčení vyplývalo, že som neomylný* vzhľadom na *p* a zároveň by som nebol neomylný* vzhľadom na *p*. Ak z mojich presvedčení vyplýva, že som neomylný* vzhľadom na *p*, tak som skutočne neomylný* vzhľadom na *p*.

Podľa bodu b) je *p* pravdivé v prípade, že z mojich konzistentných presvedčení vyplýva, že som neomylný* vzhľadom na *p* a tiež samotné *p*. Ak z mojich presvedčení vyplýva *p* a tiež to, že som neomylný vzhľadom na *p*, tak *p* je pravdivé.

Predpokladajme teraz, že moje aktuálne presvedčenia sú konzistentné a obsahujú tvrdenie, že som neomylný* vzhľadom na *p*, spolu s tvrdením *p*. Ak toto tvrdenie obsahujú, tak z nich samozrejme vyplýva. Inými slovami, práve teraz verím, že z mojich presvedčení (nech by boli akékoľvek, pokiaľ sú konzistentné), nikdy nemôže vyplývať *p* v prípade, že by bolo nepravdivé, a zároveň verím aj tomu, že *p* je pravdivé. Potom *p* je pravdivé.

Na dokázané tvrdenie sa však môžeme pozrieť aj z trocha iného uhla. Bod a) napríklad naznačuje, že žiadna množina presvedčení, z ktorej vyplýva, že daný agent je neomylný* vzhľadom na *p*, nemôže byť konzistentná, pokiaľ daný agent skutočne nie je neomylný vzhľadom na *p*. Bod b) vraví okrem iného aj to, že ak sa mýlim vzhľadom na *p* a zároveň som presvedčený, že som neomylný* vzhľadom na *p*, tak moje presvedčenia nie sú konzistentné.

Tento dôsledok bodu b) je zaujímavý, pretože sa bežne nedomnievame, že aktuálny stav sveta ($\neg p$) by mohol ovplyvniť konzistentnosť našich presvedčení ($\mathrm{B}p \wedge \mathrm{BI}p$).

4 Záver

Cieľom tohto príspevku bolo načrtnúť spôsob, akým je možné formálne skúmať pojem neomylnosti. Tento náčrt bol tiež doplnený o niekoľko výsledkov týkajúcich sa platnosti či neplatnosti rôznych formúl, ktoré reprezentujú tvrdenia o neomylnosti. Niektoré z uvedených výsledkov môžeme charakterizovať ako triviálne, no isto nie všetky.

Medzi netriviálne zistenia môžeme zaradiť napríklad tvrdenia o vzťahu medzi neomylnosťou a výrokovými spojkami (neomylnosť sa neprenáša na zložky konjunkcie, neomylnosť vzhľadom na p neimplikuje neomylnosť vzhľadom na $\neg p$). Tie poukazujú aj na jeden z možných prínosov tohto spôsobu skúmania pojmu neomylnosti: Bez dôkladnejšej formalizácie daného pojmu by možno našej pozornosti unikali.

Niektoré z prezentovaných výsledkov (najmä v časti 3.5) môžu do určitej miery budiť dojem kontroverznosti. Odporúčam ich chápať ako výsledky štúdia konzistentnosti množín presvedčení, ktoré môžu obsahovať presvedčenia o neomylnosti. Detailnejšia úvaha o predpokladaných vlastnostiach operátora možnosti M môže odhaliť, že nie sú také prekvapivé, ako sa na prvý pohľad zdá.

Uvedené výsledky sú však v každom prípade iba začiatkom a načrtnutý smer práce môže pokračovať rôznymi smermi. Zaujímavé by napríklad bolo skúmanie jazyka, v ktorom I vystupuje ako nedefinovaný operátor (i-jazyk). Keď sa zameriame na určitú triedu modelov, ako by vyzerala axiomatizácia množiny formúl (i-jazyka), platných v danej triede modelov?

Zaujímavé by bolo aj skúmanie operátora neomylnosti v kontexte iných druhov neklasických logík. Všimnime si totiž, že operátor I sme definovali v podstate pomocou pojmu striktnej implikácie[13] ($\mathrm{I}\phi$ vtt $\mathrm{B}\phi$ striktne implikuje ϕ). Inú verziu tohto operátora však môžeme definovať aj pomocou iného druhu implikačnej spojky, napríklad pomocou relevantnej implikácie. Východiskom by mohla byť práca na epistemickej interpretácii relevantných logík, pozri Bílková – Majer – Peliš – Restall (2010).

---

[13] Striktná implikácia sa definuje ako nevyhnutne pravdivá materiálna implikácia.

Dodatok: dôkazy tvrdení

V tomto dodatku uvádzame dôkazy tvrdení uvedených v hlavnom texte. Pri väčšine z nich ide o jednoduché cvičenia v modálnej logike.

Dôkaz tvrdenia 3.1. a) Vezmime si ľubovoľný základný model a nejaký svet x. Predpokladajme, že V(Bp ∧ Ip, x) = 1. To znamená, že V(Bp, x) = V(Ip, x) = 1. Podľa pravdivostných podmienok pre formulu Ip to však znamená, že V(Bp → p, y) = 1 pre všetky také y, že Rxy. Z reflexívnosti relácie R však vyplýva, že V(Bp → p, x) = 1. Platí však V(Bp, x) = 1, teda aj V(p, x) = 1.

b) Vyplýva z faktu, že každá formula tvaru Lϕ → LLϕ patrí do Log($MBel_E$) (pozri formulu (8) a nasledujúci komentár).

c) Vyplýva z faktu, že každá formula tvaru MLϕ → Lϕ patrí do Log($MBel_E$) (pozri formulu (10)).

d) Vyplýva z bodov a) a c).

Dôkaz tvrdenia 3.2. Stačí si vziať také W = {x}, v ktorom V(p, x) = 0 a V(Bp, x) = 0.

Dôkaz tvrdenia 3.3. Predpokladajme, že W = {x}, R je univerzálna na W, V(p, x) = 1 a N(x) = {W}. Jasne platí, že V(Bp → p, x) = 1, aj V(B¬p → ¬p). Teda V(Ip, x) = V(I¬p, x) = 1.

Dôkaz tvrdenia 3.4. Majme W = {x, y}, R je univerzálna na W, V(p, x) = 1, V(p, y) = 0 a N(x) = N(y) = {{y}}. Jasne platí, že V(B¬p ∧ p, x) = 1, aj V(Bp → p, x) = V(Bp → p, y) = 1. Teda V(Ip, y) = V(¬I¬p, y) = 1.

Dôkaz tvrdenia 3.5. Predpokladajme, že V(Ip, x) = V(Iq, x) = 1 pre nejaké x v nejakom M, no V(I(p ∧ q), x) = 0. To znamená, že V(Bp → p, y) = V(Bq → q) = 1 pre všetky také y, že Rxy. Na základe výrokovej logiky teda V((Bp ∧ Bq) → (p ∧ q), y) pre každé takéto y. Podľa druhého predpokladu však existuje také y, že Rxy a V(B(p ∧ q), y) = V(¬(p ∧ q), y) = 1. To by však viedlo k sporu iba vtedy, ak by sme v rámci konjunktívnych modelov mohli z V(B(p∧q), y) = 1 odvodiť V(Bp ∧ Bq, y) = 1. To však nemôžeme.

Dôkaz tvrdenia 3.6. a) Z predpokladu nepravdivosti danej formuly v nejakom x vyplýva, že existuje také y, že Rxy a zároveň V(Bp, y) = 1 a V(B(p ∧ q), y) = 0, alebo existuje také z, že Rxz a V(Bq, z) = 1 a V(B(p ∧ q), z) = 0. Tieto dôsledky však nie sú sporné.

b) Z predpokladu nepravdivosti danej formuly v nejakom x vyplýva, že existujú také y, z, že Rxy a Rxz a zároveň V(Bp ∧ ¬p, y) = V(B(p ∨ q) → (p∨q), y) =1 a V(Bq ∧ ¬q, z) = V(B(p ∨ q) → (p ∨ q), z) = 1. Z toho spor nevyplýva, stačí napríklad to, aby V(q, y) = V(p, z) = 1.

c) Z predpokladu nepravdivosti danej formuly v nejakom x vyplýva, že existuje také y, že Rxy a zároveň $V(B(p \vee q) \wedge \neg p \wedge \neg q, y) = 1$, pričom platí aj $V(Bp \rightarrow p, y) = 1$ alebo $V(Bq \rightarrow q, y) = 1$. Z týchto dôsledkov by vyplýval spor, iba ak by sme z predpokladu pravdivosti $B(p \vee q)$ v y mohli odvodiť $Bp \vee Bq$. To však jasne nemôžeme. Stačí model nastaviť tak, aby $V(Bp, y) = V(Bq, y) = 0$.

Dôkaz tvrdenia 3.7. Pozri dôkaz tvrdenia 3.5. V regulárnych modeloch platí predpoklad, ktorý potrebujeme na odvodenie sporu.

Dôkaz tvrdenia 3.8. a) Predpokladajme, že pre nejaké x máme $V(BIp, x) = 1$. To znamená, že $|Ip| \in N(x)$. Podľa vlastností extrémnych modelov teda $|\neg Ip| \notin N(x)$. Teda zároveň nie je pravda, že $ran_R(x) \subseteq |\neg Ip|$. To znamená, že existuje také y, že Rxy a $V(Ip, y) = 1$. Platí teda $V(M\,Ip, x) = 1$. Na základe bodu c) tvrdenia 3.1 dostávame $V(Ip, x) = 1$.

b) Vyplýva z bodu a) tohto tvrdenia a z bodu a) tvrdenia 3.1.

Katedra logiky a metodológie vied
Filozofická fakulta
Univerzita Komenského v Bratislave
Šafárikovo nám. 6
814 99 Bratislava
Slovenská republika
sedlar@fphil.uniba.sk

Literatúra

BÍLKOVÁ, M. – MAJER, O. – PELIŠ, M. – RESTALL, G. (2010): Relevant Agents. In: Beklemishev, L. – Goranko, V. – Shehtman, V. (eds.): *Advances in Modal Logic 2010*. London: College Publications, 22-38.

BLACKBURN, P – DE RIJKE, M. – VENEMA, Y. (2001): *Modal Logic*. Cambridge: Cambridge University Press.

BLACKBURN, P. – VAN BENTHEM, J. – WOLTER, F. (eds.) (2006): *Handbook of Modal Logic*. Amsterdam: Elsevier.

FAGIN, R. – HALPERN, J. Y. – MOSES, Y. – VARDI, M. Y. (1995): *Reasoning About Knowledge*. Cambridge (Mass.): MIT Press.

FAGIN, R. – HALPERN, J. Y. (1988): Belief, Awareness, and Limited Reasoning. *Artificial Intelligence* 34, 39-76.

HINTIKKA, J. (1962): *Knowledge and Belief. An Introduction to the Logic of the Two Notions*. Ithaca: Cornell University Press.

HUGHES, G.E. – CRESSWELL, M.J. (1996): *A New Introduction to Modal Logic*. London – New York: Routledge.

CHELLAS, B. (1980): *Modal Logic. An Introduction*. Cambridge: Cambridge University Press.

MAKINSON, D. (1965): The Paradox of the Preface. *Analysis* 25, č. 6, 205-207.
MEYER, J. -J. CH. – VAN DER HOEK, W. (1995): *Epistemic Logic for AI and Computer Science*. Cambridge: Cambridge University Press.
MEYER, J. -J. Ch. (2001): Epistemic Logic. In: Goble, L. (ed.): *The Blackwell Guide to Philosophical Logic*. Oxford: Blackwell Publishing, 183-202.
PEREGRIN, J. (2004): *Logika a logiky*. Praha: Academia.
SEDLÁR, I. (2011): V čom sa nemôžete mýliť? *Organon F* 18, č. 3, 351-362.
SZOMOLÁNYI, J. (2003): Filozofické otázky logiky V. *Organon F* 10, č. 3, 328-340.
VAN BENTHEM, J. (2010): *Modal Logic for Open Minds*. Stanford: CSLI Publications.