

EUROPA

Editori - Vladimír Kvasnicka, Jozef Kelemen, Jiří Pospíchal

MODELY MYSLE



zväzok 5

Copyright © Jan Burian, Igor Farkaš, Jana Horáková, Jozef Kelemen,
Juraj Hvorecký, Alica Kelemenová, Ladislav Kováč, Vladimír Kvasnička,
Mária Markošová, Jiří Pospíchal, Karel Pstružina, Ján Rybár, Ján Šefránek,
Radovan Šikl, Jiří Wiedermann 2008
Illustrations © Vladimír Havrilla
Cover design and layout © Bee and Honey 2008
Ako piaty zväzok edície Šimon vydalo © Vydavateľstvo Európa, s.r.o.,
Bratislava 2008

www.vydavatelstvo-europa.sk

ISBN 978-80-89111-34-3

Knihá vychádza vďaka finančnej podpore Gratex International, a.s.,
Bratislava a Slovenskej spoločnosti pre umelú inteligenciu

Obsah

Predhovor	6
Myšlení modelů <i>Jan Burian</i>	9
Konceptuálne východiská pre model stelesnenej mysle <i>Igor Farkaš</i>	35
Turingův stroj, test, tělo a jablko <i>Jana Horáková, Jozef Kelemen</i>	65
Umenie a emócie <i>Juraj Hvorecký</i>	78
Modely gramatiky života <i>Alica Kelemenová</i>	88
Bioenergetika: klíč k pochopeniu mozgu a mysle <i>Ladislav Kováč</i>	100
Kripkeho sémantický model možných svetov <i>Vladimír Kvasnička</i>	127
Modely jazyka ako dynamickej siete <i>Mária Markošová</i>	146
In silico simulácia etnických konfliktov <i>Jiří Pospíchal, Vladimír Kvasnička</i>	160
Co je to přirozený a co je to umělý <i>Karel Pstružina</i>	179
Paradoxy vizuálnej percepcie <i>Ján Rybár</i>	189
Modely usudzovania <i>Ján Šefránek</i>	200
Vnímání z pozic konstruktivismu a enaktivního přístupu <i>Radovan Šikl</i>	215
Specifikace výpočetního modelu mysli vědomého vtěleného agenta <i>Jiří Wiedermann</i>	229

Predhovor

V tomto roku si informatik, kognitívny vedec a filozof doc. Ing. Ivan M. Havel, CSc., PhD., a my spolu s ním pripomíname jeho sedemdesiatiny. Pri tejto príležitosti mu venujeme príspevky zozbierané do tejto knihy. Všetky boli napísané ako určitý druh pozdravu jubilatovi od početnosťou nepatrného zlomku jeho českých a slovenských kolegov a priateľov. Napísali ich ľudia, ktorí si vážia odbornú prácu I.M. Havla a jej výsledky, ako aj jeho všeobecnejšie názory na vedu, jej rozvíjanie a spoločenské funkcie, jeho občianske a ľudské postoje.

I.M. Havel začínal svoju odbornú dráhu ako elektroinžinier absolvovaním Fakulty elektrotechniky pražskej ČVUT. Už vo svojej diplomovej práci sa zamerl na vtedy veľmi aktuálnu a pomerne ešte novú oblasť, ktorá s elektrotechnikou síce súvisela, nasmerovala však jeho ďalšie odborné záujmy úplne inam, než k elektrotechnike. Tou oblasťou bola teória automatov. Od nej sa dostal do širšej oblasti vedy, do teoretickej informatiky. V rámci svojho doktorandského štúdia, ktoré absolvoval na Kalifornskej univerzite v Berkeley, sa venoval striktno deterministickým jazykom, výsledky tohto svojho mimoriadne plodného obdobia publikoval v sérii prác so skvelou medzinárodnou odozvou. Už v dobe týchto úspechov sa však chystal na ďalšiu zmenu svojho odborného záujmu. Tú predznamenovali jeho návštevy prednášok z robotiky a umelej inteligencie na Kalifornskej univerzite, na ktorej získal titul PhD. Zúročiac skúsenosti z teoretickej informatiky a to, čo vedel z matematickej logiky, dosiahol aj v oblasti umelej inteligencie medzinárodne akceptované a citované výsledky.

Po takejto skúsenosti, podobne ako mnohí iní, ani on neodolal pokušeniu venovať sa niektorým všeobecnejším problémom myslenia, inteligencie, vnímania, prežívania, vedomia... Tieto otázky dodnes zamestnávajú tých, ktorí skúmajú ešte stále dosť neznámu oblasť na pomedzí umelej inteligencie, psychológie, filozofie mysle, neurovied a niektorých ďalších oblastí dobrodružnej ľudskej zvedavosti – kognitívnu vedu. Jeho kniha „*Robotika - úvod do teorie kognitivních robotů*“ vydaná v r. 1980 bola v Československu prvou monografiou, ktorá systematickým spôsobom prezentovala symbolické metódy umelej inteligencie. Väčšina z nás, ktorí sa vtedy začínali zaujímať o tento informatický vedný odbor, môže konštatovať, že práve táto kniha bola pre nás primárnym zdrojom vedomostí o umelej inteligencii (a snáď aj o kognitívnej vede, ale vtedy sme ani len netušili, že takýto vedný odbor silne interdisciplinárneho charakteru práve v USA vzniká).

Jeho prednáška „*Neuronové počítače a jejich inteligence*“, ktorú prezentoval na konferencii SOFSEM v r. 1988, bola pre väčšinu nás

„bleskom z jasného neba“, kde sme sa oboznámili so základnými ideami opätovne vznikajúcej teórie neurónových sietí a so subsymbolickým prístupom k riešeniu problémov kognície. Od neurónových sietí a kognitívnej vedy je iba maličký krôčik k filozofii mysle. Ivan M. Havel sa bez rozpakov odhodlal vykonať ho. Tieto snahy našli určité vyvrcholenie publikovaním rozsiahlej state „*Přirozené a umělé myšlení jako filosofický problém*“, ktorá vyšla v r. 2001 v zborníku „*Umělá inteligence III*“, pravidelne vydávaného V. Maříkom a spol.

Záujmovými posunmi, ktoré dávali smer a určovali charakter jeho mnohostranného odborného profilu, dosiahol Ivan M. Havel okrem iného to, že má dnes kolegov, ktorí s úctou hľadajú na jeho cestu vedúcu a príslušnými oblasťami premýšľania, i na výsledky, ku ktorým ho jeho cesta postupne privádzala. Spomedzi týchto kolegov sa grupujú autori príspevkov napísaných pre tento zväzok. Čo ich spája, je využívanie príležitostí stretávania sa už po sedem rokov. Tieto príležitosti existujú vďaka hŕstke českých a slovenských nadšencov a sú venované veľmi široko chápanej kognícii, hlavne však kognitívnej vede, filozofii mysle a problematike umelého života. Sú to stretnutia *Kognícia a umelý život*. Dúfame, že textami, ktorými zdravíme jubilanta, mu privodíme radosť z vedomia priateľstva prispievateľov i intelektuálne potešenia z čítania ich textov. Vydanie knihy podporili *Slovenská spoločnosť pre umelú inteligenciu* a *Gratex International, a. s.* Ďakujeme im, bez ich podpory by kniha nevyšla. Ďakujeme vydavateľovi, bez starostlivosti ktorého by kniha nebola taká, aká je. Veľmi si však prajeme aj to, aby dobrý pocit z čítania mali nielen jubilant, ktorému ju týmto venujeme, ale i všetci ostatní, ktorí ju poctia svojím záujmom. Bez toho by kniha vyšla zbytočne.

V Bratislave v marci 2008

editori



Myšlení modelů

Jan Burian¹

Abstrakt. Cílem tohoto článku je metodologický rozbor předpokladů, které používáme, když přemýšlíme o pojmech jako mysl, myšlení, vědomí, poznávání a zkušenost. V první části ukážeme, za jakých předpokladů má smysl chápat mysl jako proces, který je možno popsat formálním modelem a kde jsou meze modelování. V druhé části se zaměříme na alternativní metodologické přístupy, které umožňují poznávat mysl, resp. zkušenost nikoliv pomocí formálních modelů, ale reflexivním nepojmovým poznáváním, které budeme označovat jako vhléd.

V kognitivní vědě hraje sebereflexe a subjektivita dosti pozoruhodnou roli. Na jedné straně je skryta v pozadí téměř všeho, o čem se v kognitivních vědách mluví – skryta proto, že badatelé ve snaze vyhovět požadavku objektivní poznání, který je jedním z hlavních postulatů moderní přírodovědy, neradi zdůrazňují intuitivní původ svých pojmů, problémů a hypotéz. Na druhé straně jsou to právě kognitivní vědy – a doložitelně umělá inteligence – které nastolily problém já jako relevantní výzkumné téma.

Ivan M. Havel (2001)

1. Úvod

Cílem tohoto článku je metodologický rozbor předpokladů, které používáme, když přemýšlíme o pojmech jako mysl, myšlení, vědomí, poznávání a zkušenost. Konkrétně ukážeme, za jakých předpokladů má smysl chápat mysl jako proces, který je možno popsat formálním modelem a kde jsou meze modelování. V první části, v kapitole o tradičních modelech mysli, se soustředíme na formální analýzu poznávacích procesů z hlediska 3. osoby, tj. budeme popisovat poznávací procesy z hlediska vědce, který se snaží o objektivní popis. Tato kapitola bude zajímavá zejména pro odborníky na umělou inteligenci. V druhé části článku se soustředíme na poznávání z hlediska 1. osoby, tj. na reflexivní zkušenost. Zaměříme se zejména na alternativní metodologické přístupy, které umožňují poznávat mysl, resp. zkušenost, nikoliv pomocí formálních modelů, ale nepojmovým poznáváním, které budeme označovat jako vhléd. Tato kapitola bude zajímavá zejména pro filosofy, ale i pro psychoterapeuty a pedagogy.

¹ Vysoká škola ekonomická, Fakulta národohospodářská, Katedra filosofie, nám. W. Churchila 4, 190 00, Praha 3, e-mail: burianj@vse.cz

1.1 Pojmy

Nejprve definujeme pojmy, které budeme používat.

Vyjdeme z pojmu zkušenost². Zkušeností (experience) nebudeme chápat pouze zapamatovanou znalost či dovednost, ale celé spektrum aktuálního zakoušení nebo prožívání – jde o veškeré poznávací procesy jako je smyslové vnímání, pojmové myšlení, pocity a emoce, chtění a záměry a další (sebe)reflexivní procesy jako uvědomování, zaměřování a udržování pozornosti apod. Aktuální zkušenost je ale vždy ovlivňována zkušeností minulou. Pozorujeme, že zkušenost se většinou řídí nějakými více či méně stabilními pravidly, která vznikla interakcí vrozených dispozic a v průběhu života získaných návyků. Nebudeme však předpokládat, že tyto zákonitosti se nemohou měnit. Zkušenost tedy zůstává otevřená k libovolně hluboké proměně. Proměnu zkušenosti dosahovanou pomocí reflexe, tj. pomocí pozorování sebe sama, budeme nazývat transformace. Nebudeme dále předpokládat, že existují nějaké předem dané meze normální zkušenosti, a že překročení těchto mezí musí nutně znamenat zkušenost patologickou.

Pojem vědomí (consciousness) má specifický význam v medicíně, kde označuje stav pacienta, který dokáže reagovat na podněty jinak než čistě reflexivně, anebo i když schopen reakce není (např. při vážném ochrnutí), je schopen si podněty zapamatovat a později re-produkovat. Ve filosofii označuje vědomí v podstatě zkušenost s důrazem na subjektivní prožívání v 1. osobě. Vědomí má kvalitativní a intimní, obtížně sdělitelnou povahu. V materialistickém či fyzikalistickém kontextu (viz níže metafyzický realismus) existuje tendence postulovat kauzální závislost vědomí na objektivně existujícím vnějším (external) fyzickém světě. Fenomenologicky založení myslitelé mají spíše tendenci chápat fyzický svět jako druh vědomé zkušenosti a fyzikální zákony někdy chápou jako odvozené od zákonitostí smyslového poznávání.

Mysl (mind) je pojem, který nemá jasnou definici, resp. má mnoho různých definic. Všeobecně jako mysl chápeme souhrn a podklad poznávacích procesů. Někdy je tento termín používán jako synonymum k pojmu vědomí, pak bývá často zdůrazňována subjektivní povaha mysli a problém jejího vztahu k objektivnímu fyzickému tělu (mind-body problem). Někdy je význam pojmu mysl širší než vědomí a zahrnuje i nevědomé či podvědomé procesy. Tradiční kognitivní věda, chápe mysl velmi specificky jako soubor procesů pojmového (konceptuálního) myšlení, kde pojmy jsou reprezentace objektivního

² V (Burian, 2004) a (Burian, 2007) jsem ve stejném významu, jako zde používám pojem zkušenost, používal pojem poznávání.

světa, které je možno označit symboly a pojmové myšlení pak redukovat na logické operace se symboly.

Pojem vědomí i mysl jsou zatížené řadou různých interpretací a v tomto článku je budeme používat pouze v kontextu teorií, které s těmito pojmy pracují. Ve zbytku článku budeme pracovat s pojmem zkušenost. Často budeme také pracovat s pojmem poznávání (cognition). Význam tohoto termínu v rámci různých paradigmat kognitivní vědy probereme blíže v kapitole o tradičních modelech mysli.

1.2 Změna paradigmatu

Věda je proces, na kterém participují vědci. Vědci jsou poznávající agenti. Vědci shromažďují údaje vzniklé na základě pozorování přirozených či experimentálně iniciovaných dějů. Vědci tvoří teorie, ze kterých se dají vyvodit potenciálně vyvratitelné předpovědi. Vědci vzájemně diskutují (posuzují, ověřují) výsledky pozorování i teorie. Pozorování, shromažďování a zpracování údajů, tvorba teorií i vědecký diskurz jsou pro každého vědce složky zkušenosti. Věda je tedy pro každého vědce zkoumáním zkušenosti.

Naprostá většina naší běžné každodenní zkušenosti je slučitelná s předpoklady tzv. metafyzického realismu. Metafyzický realismus předpokládá, že existuje objektivní, na zkušenosti nezávislá, realita³, a že zkušenost, zejména pokud je korigována vzájemným diskurzem, je relativně přesnou, i když ne třeba úplnou, reprezentací objektivní reality.

Objektivní realita je tvořena fyzikálními strukturami a zákonitostmi jejich vývoje. Tyto struktury a zákonitosti jsou poznatelné pomocí pozorování a to je převoditelné na pojmové myšlení, které je zase formálně popsitelné a dá se (alespoň v principu) modelovat pomocí počítače. Metafyzický realismus je v souladu s lidovou psychologií mající kořeny v západním (euro-americkém) světonázoru. Toto metafyzické paradigma je zároveň v souladu s většinou vědeckého výzkumu, zejména nijak nebrání pragmatické aplikaci získaného poznání do konkrétních technologií.

Historie vědy však ukazuje, že velké změny paradigmatu nastávají právě v situaci, kdy se zdá, že jsou všechny fundamentální problémy vyřešeny a zbývá jen několik „drobností“, které je třeba „dovysvětlit“. Podobně tomu bylo i na přelomu 19. a 20. století. Newtonovskou fyziku, která uspokojivě vysvětlovala naprostou většinu běžně pozorovaných jevů, bylo nutno zasadit do širšího rámce teorie relativity

³ Existuje řada objektivistických metafyzik idealistických, které však nepostulují objektivní realitu, tj. věcnost, ale nějakou objektivní podstatu duchovní – Bůh, Já, absolutní duch, prvotní hybatel. Idea Dobra, átman apod.

a kvantové mechaniky, aby bylo možno vysvětlit takové jevy jako je chování černého tělesa či nepřekročitelná rychlost světla. Pro současnou vědu by takovým problémem mohla být existence procesů, které nelze formálně popsat, modelovat pomocí počítače a pochopit pojmovým myšlením.

2 Tradiční modely mysli a jejich omezení

Každá historická epocha používala k popisu fungování mysli více či méně přílehlavé metafory, většinou odvozené od v té době nejsložitějších strojů. Mysl tak byla v historii přirovnávána např. ke katapultu, mlýnu, hodinovému mechanismu, hydraulickému systému, telegrafní či telefonní centrále a nejnověji k počítači. Přestože počítačová metafora je velmi praktická, je možné, že v budoucnosti bude tato metafora stejně úsměvná jako je dnes metafora katapultu či mlýnu.

Připomeňme v krátkosti jaké modely mysli, resp. poznávacích procesů, má současná kognitivní věda k dispozici⁴.

2.1 Kognitivistické a konekcionistické paradigma

Matematici a logici jako Alan Turing, John von Neumann či Alonzo Church, kteří ve 30. a 40. letech navrhovali první formální modely počítačů, se inspirovali vlastní zkušeností s řešením logických problémů. Navrhovali modely, které napodobovali způsob jakým oni řešili logické problémy. Proto jsou počítače, tj. zařízení výpočetně ekvivalentní univerzálnímu Turingovu stroji (UTM)⁵, vlastně stroje na automatické operace se sekvencemi symbolů.

Jaká jsou omezení počítačové metafory a současné vědeckého popisu vůbec?

Každý popis je z formálního hlediska sekvence symbolů. Symboly jsou vybírány z nějaké abecedy, které má minimálně dva prvky, např. 0 a 1. Pokud si věda klade ambice popsat tu složku zkušenosti, kterou nazýváme fyzikální realita, pak tím tvrdí, že fyzikální realita je převoditelná na sekvence symbolů. Nemáme ovšem zaručeno, že existuje algoritmus, který dokáže generovat všechny takové sekvence, které popisují stavy fyzikální reality. Teprve pokud by takový algoritmus existoval, znamenalo by to, že fyzikální realita je tzv. spočetná, tj. v principu plně modelovatelná počítačem.

Vědecký popis má většinou formu teorie. Teorie je soubor formulí, což jsou takové sekvence symbolů, kterých je možno určit logickou pravdivostní hodnotu. K teorii musí být přiřazen soubor logických pravidel

⁴ Podrobněji je z hlediska teorie informace toto téma zpracováno v (Burian, 2005a).

⁵ Samozřejmě u reálných počítačů nepředpokládáme, na rozdíl od Turingova stroje, nekonečnou paměť.

určujících, jakým způsobem budou symboly tvořící stávající formule převedeny na nové formule (proces logického odvozování, inference). Teorie by měly mít některé základní vlastnosti, zejména by měly být logicky konzistentní, tj. z předpokladů by měly jít pomocí pravidel logické inference odvodit pouze závěry, které nejsou logicky sporné. Tyto závěry by dále měly být interpretovatelné jako předpovědi, které jsou potenciálně vyvratitelné a dají se testovat pomocí pozorování a experimentů. Model je takové ohodnocení proměnných nějaké teorie, při kterém každá formule této teorie bude pravdivá.

Pokud je tedy teorií to, že mysl je možno modelovat pomocí Turingova stroje, teorií myslí je definice Turingova stroje. Modelem jsou pak konkrétní vstupní a výstupní abeceda, množina stavů a přechodová pravidla, která budou převádět počáteční sekvenci symbolů na pásce (nebo množinu přípustných sekvencí) na koncové sekvence symbolů. Chování stroje by mělo v nějaké interpretaci odpovídat buď empirickému pozorování v případě modelů přirozených organismů, nebo nějakému požadovanému výsledku v případě prakticky zaměřených technologií.

Tento přístup k modelování poznávacích procesů byl přímou inspirací tzv. kognitivistickému paradigmatu kognitivní vědy. V tomto paradigmatu jsou symboly přiřaditelné pojmům. Pojmy jsou pak reprezentace konkrétních (extenze) či obecných (intenze) prvků objektivní reality a vztahů mezi nimi. Pojmy můžeme formálně interpretovat jako pozorovatelné pravděpodobnostní veličiny, které mohou nabývat různých hodnot a definují stavový prostor systému. Vztahy mezi pojmy můžeme chápat jako statistické závislosti, a k popisu operací se symboly pak můžeme použít teorii informace.

Konekcionistické paradigma je někdy stavěno do ostrého protikladu k paradigmatu kognitivistickému. Konekcionismus je inspirován výzkumem fyziologie poznávacích procesů v živých organismech. Konekcionismus modeluje poznávání jako dynamický vzorec interakcí ve složité síti složené z relativně jednoduchých prvků a jejich vztahů (např. neuronová síť). Prvky této sítě lze chápat jako symboly, ale těmto symbolům již nelze přiřadit konkrétní pojmy reprezentující objekty⁶. Jako reprezentaci objektivní reality je však možno chápat celkový stav konekcionistického systému.

Konekcionismus a kognitivismus však k sobě mají blízko, protože i paralelní interakce prvků v konekcionistické síti lze modelovat jediným klasickým počítačem⁷.

⁶ Konekcionistické paradigma bývá někdy označováno také jako subsymbolické paradigma. To je v pořádku pokud ztotožňujeme symboly a pojmy (koncepty), v tomto textu ale symboly a pojmy rozlišujeme, takže správnější by bylo označení subkonceptuální paradigma.

V obou těchto paradigmatech lze dokonce modelovat i mysl přemýšlející o modelech myšlení. Stačí si uvědomit, že univerzální Turingův stroj, může emulovat jiné univerzální Turingovy stroje. Co však možné není, je dokonalé modelování přímé reflexe myšlení, kdy mysl přemýšlí sama o sobě. To by odpovídalo situaci, kdy na vstupu UTM je jeho vlastní kód, který má na vstupu svůj vlastní kód atd. a došlo by k zacyklení. Ze zkušenosti víme, že dokážeme více či méně dokonale pojmově přemýšlet o svém myšlení. Na to by se dalo namítnout, že pojmová reflexe nikdy není dokonalá, nikdy nenazíráme dokonale přesně, že přemýšlíme o tom jak myslíme, že myslíme atd. Na vstupu onoho UTM modelujícího reflektující mysl by tedy mohly být jen zjednodušené či částečné modely tohoto UTM.

Závažnější problém je, zda pojmová myšlení vůbec může být formou reflexe, zda se nejedná pouze o konstrukci pojmových struktur. Ze zkušenosti víme, že o procesech probíhajících v naší zkušenosti nemůžeme pojmově přemýšlet, ale můžeme je prostě jen pozorovat. V dalších kapitolách budeme řešit právě otázku, zda je možné toto reflexivní nepojmové pozorování formálně modelovat a pokud nikoliv, zda může přinášet kvalitativně jiné poznatky než pojmové myšlení.

2.2 Enaktivní přístup – hledisko 3. osoby

Jako třetí paradigma kognitivní věd bývá někdy označován tzv. enaktivní (zjednávací⁸) přístup (Varela et. al., 1991). Tento přístup vychází z konstruktivistických proudů v kybernetice (součást širokého proudu tzv. second-order cybernetics). V tomto přístupu není poznávání chápáno jako pasivní reprezentace objektivního světa, ale jako aktivní proces interakce mezi poznávajícím agentem (např. organismus, robot apod.) a poznávaným prostředím. Agent je charakterizován jako autonomní (řídí sám sebe), situovaný (vždy v kontextu nějakého prostředí) a vtělený (embodiment, agent má tělo, které mu umožňuje interakci s prostředím). V různých kontextech bývají zdůrazňovány odlišné charakteristiky my se zaměříme zejména na vtělenost⁹.

Enaktivní poznávání můžeme chápat ve dvou odlišných kontextech. Jednak můžeme interakci mezi poznávajícím agentem a prostředím popisovat v roli pozorovatele (observer) ve 3. osobě. Např. vědec pozoruje pomocí přístrojů poznávací procesy u nějakého organismu či umělého agenta. Nebo poznávajícím agentem může být sám

⁷ Model konekcionistického systému běžící na jednom počítači bude samozřejmě mnohem pomalejší, než kdyby každý prvek tohoto systému představoval jeden počítač.

⁸ Autorem českého překladu pojmu enaction jako zjednávání je I. M. Havel.

⁹ Jiným komplementárním přístupem je např. rozbor možností jak chápat kognitivní procesy u autonomních agentů jako samoorganizaci (Burian, 2006).

pozorovatel v 1.osobě. Laskavý/á čtenář/ka tohoto textu, může např. právě nyní zkusit pozorovat způsob, jakým jeho oči sledují text. Zjistí patrně, že oči se nepohybují plynule, ale trhaně, že nečte jednotlivá písmena, ale celá slova, či sousloví atd.

Ve 3. i v 1. osobě jde o zkušenost pozorovatele. V prvním případě jde o zkušenost pozorování poznávacího procesu, v druhém o zkušenost reflexe poznávacího procesu. Mezi perspektivou 1. a 3. osoby je možná interakce (viz např. Varela, 1996), avšak pokud nejsou tyto dva kontexty jasně odlišeny dochází k omylům a nejasnostem.

V této kapitole se budeme zabývat enaktivním poznáváním ve 3. osobě, ve 3. kapitole pak reflexivním poznáváním v 1. osobě.

Z hlediska enaktivního přístupu poznávací procesy aktivně působí na prostředí, tj. pro pozorovaného agenta neexistuje žádná objektivní realita nezávislá na jeho poznávání¹⁰. Avšak z hlediska pozorovatele – vědce je stále možné objektivní realitu nezávislou na jeho zkušenosti předpokládat

Hlavní inspirací enaktivního přístupu je výzkum adaptace agenta na prostředí. V praxi není možné analyticky určit všechny stavy, které mohou efektivně řešit adaptaci na neustále zpětnovazebně se měnící prostředí, ať už reálné nebo i složité virtuální. Jedinou cestou je nechat agenta s prostředím interagovat, „žít“ v prostředí a adaptovat se na základě této interakce.

Vtělení se nemusí vylučovat s kognitivistickým a konekcionistickým přístupem. V první řadě je možné enaktivní přístup aplikovat i k popisu interakce počítačem simulovaného agenta v simulovaném prostředí. Jde většinou o různé multiagentní modely, nebo o softwarové multiagentní systémy. Ale i většina současných aplikací vtělených systémů v reálném prostředí je kombinací enaktivního přístupu

¹⁰ Výsledek poznávání musí být někde uložen, takže i když předpokládáme, že se nemění poznávané objekty (tj. poznávané části objektivní reality), musí se v důsledku poznávání, skrze řadu interakcí, měnit objektivní realita někde jinde. Tuto změnu označíme jako data o poznávaném objektu. Použití pojmu informace v tomto kontextu by bylo nepřesné (viz Burian, 2005a)

Pokud bychom chtěli trvat na předpokladu, že poznávání nemění objekty, pak musíme předpokládat, že objekty, jsou navzájem nezávislé, že změna způsobená uložením dat o poznávaném objektu nebude na tento objekt zpětně působit. Dá se ukázat, že tento předpoklad, je principiálně nesplnitelný, že objekty od sebe nelze absolutně izolovat. Ve většině praktických případů, je však vliv ukládání dat o poznávaném objektu na poznávaný objekt tak malý, že se dá zanedbat. Výjimkou je poznávání mikroskopických objektů, kde objekty, kterými je poznávání (měření) prováděno, jsou opět mikroskopické objekty. Interakce mezi nimi, která vede k uložením dat o poznávaném, je natolik silná, že zanedbatelně mění stav poznávaného mikroskopického objektu.

Ovšem i v makroskopickém světě můžeme zkonstruovat podobný jev – představme si, že máme se zavázanýma očima rozpoznat obrazec načrtnutý před námi v jemném písku. Pomocí hmatu sice poznáváme zhruba o jaký obrazec jde, ale zároveň původní podobu obrazce narušujeme.

a kognitivismu či konekcionismu. Ve skutečném prostředí se pak pracuje zejména se senzomotorickou koordinací. Roboti řízení klasickým počítačem nebo umělou neuronovou sítí (často simulovanou klasickým počítačem) se učí pohybovat ve složitém terénu, komunikovat s dalšími roboty, apod. Enaktivní přístup však neznamená, že pouze připojíme kognitivního agenta na jedné straně k sensorům, na druhé k motorice a navrhne systém jejich vazeb, byť sofistikovaný, ale v zásadě pevně daný.

Míra adaptace agenta na prostředí závisí na míře vtělení agenta. Míra vtělení výše zmíněných systémů je omezená, neboť počítačem řízené systémy jsou vždy omezeny nějakými základními prvky designu, základní sadou instrukcí, apod. U formálních modelů jako je Turingův stroj je tímto omezením nutnost určit pevně danou abecedu symbolů, stavy, přechodové funkce atd. U agentů fungujících ve skutečném prostředí jsou tato omezení realizována ve formě hardware. Takový agent je tedy vtělen jen do jisté míry, vždy existuje nějaká pevná hranice mezi tělem a reprezentací prostředí. Není možné plně ztotožnit agenta s jeho tělem. Na úrovni základních prvků existuje tedy pevně definovaná hranice, která odděluje agenta od prostředí. Toto omezené vtělení budeme označovat jako uzavřené.

Pokud myšlenku vtělenosti poznávání dovedeme do důsledků, poznávající agent může být vtělen do té míry plně, že se stává součástí prostředí, které proměňuje. Může tedy proměňovat i sám sebe. Taková míra vtělení znamená, že nelze oddělovat působení prostředí na agenta např. ve formě sensorických vstupů a aktivního působení agenta na prostředí např. ve formě motorické. Interakce mezi agentem a prostředím je vždy zároveň i aktivním sebemodifikací či sebevytvářením (autopoiesis – viz např. Maturana, 2002; Maturana & Varela 1987). Tato sebemodifikace přitom není principiálně nijak omezena, může se rozvíjet například (vtěleným) evolučním procesem. V takovém případě je možno v limitě ztotožnit agenta a jeho tělo. Systém totiž může rozvíjet i svoji schopnost ovlivňovat prostředí a tím schopnost sebemodifikace zdokonalovat. Proto není předem určeno, jaké jsou hranice poznávajícího systému, respektive co vlastně je poznávající systém a co už je prostředí. Plná vtělenost systému umožňuje opustit předpoklad existence pevně definovaných hranic systému. Toto plné vtělení budeme označovat jako otevřené. Poznamenejme, že uzavřenost a otevřenost u vtělených agentů znamená něco jiného než uzavřenost a otevřenost systémů v termodynamice.

2.3 Otevřené agenty

*Zkrátka své životy nemůžeme vypočítat – my je žijeme.
A pomocí příběhů přikládáme našim činům závislým
na kontextu smysl. My autonomní agenti.*

Stuart Kauffman (2004)

Mohou existovat simulované otevřené agenty v simulovaném prostředí – simulované prostředí je sice vždy nějakým způsobem uzavřeno, ale interakce agenta s prostředím může být v rámci těchto omezení navržena jako otevřená, tj. agent je plně součástí prostředí a plně v něm modifikovatelný. Příkladem takových systémů jsou některé modely umělého života jako Tierra, kde každá část prostředí může být alespoň potenciálně částí těla agenta a těla agentů tvoří prostředí pro sebe navzájem. V dalším textu budeme ale pod pojmem otevřený agent rozumět otevřené vtělené agenty v reálném prostředí.

V současnosti však neexistuje žádný umělý otevřený agent. Přesto si dokážeme do jisté míry takového agenta představit. Nejednodušší formou sebemodifikace je sebereprodukce spojená s jistou mírou mutací. Už John von Neumann se pokoušel v 60. letech navrhnout vesmírnou sondu, která by z nalezeného materiálu stavěla vlastní kopie. Přestože plně sebereprodukující umělý agent nebyl dosud zkonstruován, je jistě v principu možné ho sestavit a pravděpodobně k tomu dojde v blízké budoucnosti. Existují dva základní směry jak takové agenty konstruovat. První je klasický – agent bude mít podobu robota řízeného počítačem. Druhý směr je zajímavější, je inspirován potřebou biotechnologického průmyslu automatizovat výrobu složitých molekul. Objevuje se poptávka po nanotechnologických zařízeních o velikostech virů nebo buněk, které budou schopny skládat (assemble) složité molekuly. Aby však tato zařízení dosahovala rozumné produktivity, je třeba jich nejprve zkonstruovat obrovské množství. Jediný efektivní způsob je patrně sebereprodukce.

Přestože jsme se zatím bavili pouze o sebereprodukci nikoliv o mutaci, je zřejmé, že není zásadní problém umožnit při reprodukci jistou míru volnosti v konstrukci nového agenta. Na molekulární úrovni se patrně mutacím nebude ani možno vyhnout. Mutace spolu s nějakým typem selekce (ať už přirozené či umělé) povede k evoluci.

Všechny lidské technologie mají daný nějaký konstrukční plán, který umožňuje pojmově zachytit charakteristiky podstatné pro fungování této technologie. Každá technologie má na začátku daný nějaký účel, k něčemu ji chceme používat. Schopnost plnit tento účel

určuje, zda jsme s technologií spokojeni či nikoliv. Pokud nikoliv, snažíme se ji nahradit jinou technologií. Schopnost plnit účel tvoří další pojmově uchopitelnou dimenzi, kterou můžeme chápat analogicky k fitness (míra přizpůsobení) živého organismu. Avšak živé organismy nemají žádný předem daný účel kromě adaptace a přežití, a proto se i charakteristiky, které určují jejich fitness mohou měnit. Molekulární sebemodifikující se agenty, přestože jejich vznik byl iniciován uměle, se v podstatě začínají podobat živým organismům. Zejména pokud se vymknou kontrole, pokud přestanou plnit lidmi určené účely. Dokonce je možné spekulovat o tom, že tyto agenty mohou mít (nebo v průběhu evoluce získat) ještě vyšší sebemodifikační schopnosti než živé organismy vzniklé přirozenou evolucí a mohou se tak s jistou nadšázkou stát „živější než živé“.

Živé organismy, pokud je chápeme v evoluční perspektivě, jsou příklady otevřených agentů¹¹. Stuart Kauffman (2004) spekuluje o tom, že nelze stanovit stavový prostor biosféry, tj. stavový prostor obsahující všechny možné trajektorie evoluce. Jedním z důvodů, které udává je existence tzv. exaptací. Exaptace je vlastnost organismu nebo i umělého systému, která byla sekundárním důsledkem nějaké adaptivní změny a neznamena sama o sobě pro organismus žádnou výhodu, nezvyšovala jeho fitness, mohla být dokonce i jistou nevýhodou. V novém kontextu, např. pokud organismus přišel do nového prostředí, mohla se tato vlastnost stát konkurenční výhodou a dále se posilovat pomocí přírodního výběru. Obecně se dá říci, že exaptace mohou měnit pravidla, podle kterých aktuálně evoluce funguje. Kauffman uvádí řadu příkladů, z nichž stručně popíšeme pouze jeden. Jde o příběh evoluce létající veverka (poletuchy) – zvířátka, které žije podobně jako veverka na stromech, ale vyvinula se jí volná kůže na bocích, která jí umožňuje překonávat plachtěním velké vzdálenosti. To je samozřejmě výhoda, avšak předtím než se volná kůže na bocích stala výhodou, pravděpodobně předcházela fáze, kdy tato kůže spíše překážela nebo neměla žádnou funkci.

¹¹ Jako další příklad uvedme hru NOMIC od Petera Subera (1990). NOMIC je hra imitující vznik legislativy. Na začátku máme sadu pravidel, které určují jak hrát nějakou hru, ale zároveň určují, jak měnit tuto sadu pravidel. Pravidla je možné měnit zcela libovolně – na začátku je určeno, že se rozhoduje demokraticky, ale to se může změnit. Stejně tak se může změnit cíl hry, nebo může být hra zcela zrušena. Může být dokonce zrušeno i první pravidlo – že je nutno se řídit dle pravidel – hráči pak ovšem dostanou co chtějí. Přestože na první pohled vypadá tato hra jako formální systém pravidel, jejich sémantika vyžaduje lidské porozumění i aktivitu – tj. otevřeně vtělení ve skutečném prostředí. Mohou totiž existovat např. pravidla typu: Jen pravidla napsaná na kusu papíru jsou platná. Kdo spálí papír s pravidly vyhrál. Pokusy formalizovat NOMIC jsou v plenkách a ukazují, že je to do jisté míry možné, pouze pokud radikálně omezíme množinu možných typů pravidel. Jenže to pak již není původní zcela otevřený NOMIC.

Existence exaptací znemožňuje předem zjistit, které pojmově uchopitelné vlastnosti organismu a prostředí jsou podstatné pro fitness organismu. Pokud neznáme tyto pojmy nemůžeme ani určit veličiny, které by tvořili dimenze stavového prostoru evoluce a tedy není možné vytvořit přesný model evoluce. Můžeme samozřejmě vytvářet více či méně zjednodušené modely evoluce, ty však nikdy nemohou předpovědět veškeré bohatství možných evolučních trajektorií.

Můžeme samozřejmě namítnout, že pokud můžeme modelovat fyzikální realitu pomocí fyzikálních zákonů, pak je možné v principu modelovat i evoluci biosféry, která je součástí fyzikální reality.

Zde vyvstává několik problémů. Kauffman se snaží ukázat, že vývoj vesmíru a biosféry v něm je silně neergodický (tj. vývoj vesmíru a biosféry nebude opakovat minulé stavy) a prostor všech potenciálně možných stavů biosféry expanduje tak rychle a je tak obrovský, že i kdybychom měli k dispozici další vesmír jenom na modelování vývoje biosféry, tak bychom dokázali prozkoumat jen zcela nepatrný zlomek tohoto prostoru.

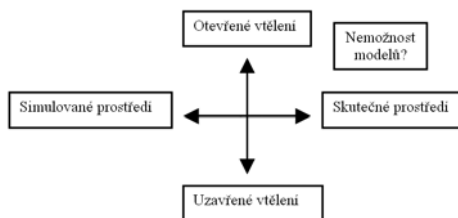
Další problém je, že pokud bychom modelování prováděli na kvantové úrovni, nikdy bychom nemodelovali aktuální stavy biosféry, ale pouze pravděpodobnostní rozložení možných stavů.

Kauffman však jde ještě dál a uvažuje o možnosti, že i fyzikální zákony nejsou neměnné, ale může se jednat pouze o relativně stabilní fáze ve vývoji vesmíru. Fyzikální zákony by se mohly měnit podobným procesem jako evoluce organismů, stabilní by zůstávali vždy ty zákony, které jsou nejlépe „přizpůsobené“ aktuálnímu stavu vesmíru¹². Pokud bychom Kauffmanovy inspirace dále rozvedli, mohli bychom vesmír začít chápat jako otevřený vtělený agent. Není předem známo, zda existují nějaké „metazákony“, podle kterých se budou fyzikální zákony měnit, jaká jsou kritéria pro stabilizaci zákonů, jaké pojmy použít pro popis změn zákonů. I kdybychom se pokusili takové „metazákony“ odhalit, zůstává otázka, zda i ony nepodléhají změnám. V takovém případě by byla fyzikální realita z principu nevyčíslitelná a nemodelovatelná. Tj. není možné zkonstruovat seznam jejích stavů – protože k tomu potřebuji neměnnou sadu základních atributů (jako jsou dimenze prostoru, času, hybnosti, energie, základních sil), které jsou konceptuálně uchopitelné a mohu je formálně označit symboly. Přitom je z praktického hlediska jedno, zda tyto „metazákony“ objektivně existují, jenom je není možno předem zjistit, nebo zda neexistují a změny zákonů můžeme našimi omezenými prostředky popsat pouze jako „náhodné“.

¹² Podobnou představu, ovšem argumentovanou z filosofických pozic, je možno najít např. v (Burian, 2004).

Uzavřeme tyto poněkud „divoké“ úvahy diagramem, který ukazuje dvě dimenze, podle kterých je možné klasifikovat poznávající agenty z hlediska enaktivního přístupu.

Horizontální dimenze představuje škálu od plně simulovaného virtuálního prostředí, které je jednoznačně vyčíslitelné, až po skutečné prostředí. Vertikální dimenze představuje míru vtělení agenta od uzavřeného po otevřené vtělení. Z hlediska vyčíslitelnosti a modelovatelnosti je problematický pravý horní kvadrant do něhož patří živé organismy.



Obr. 1. Dimenze pro klasifikaci enaktivních agentů

Můžeme zformulovat hypotézu, že část naší zkušenosti pozorování poznávacích procesů (tj., enaktivní poznávání ve 3.osobě) nelze formálně popsat a modelovat. Konkrétně jde o pozorování otevřených agentů. Tato hypotéza a její důsledky neohrožují sice samotný předpoklad, že existuje objektivní realita nezávislá na zkušenosti pozorovatele, ale narušuje důležitý argument podporující možnost objektivní realitu správně poznávat. Nemožnost formálního popisu vede totiž k nemožnosti úplné komunikace zkušenosti pozorování mezi pozorovateli - vědci, tj. je překážkou vzájemné validace pozorování pomocí diskurzu. Přestože je možná komunikace pomocí zjednodušených modelů, není možné vyloučit, že vědci budou mít značně odlišné zkušenosti pozorování otevřených agentů.

3 Zkušenost a metafyzika

I wish to insist in that the epistemological shift in the notions of autopoiesis and the biology of cognition that I have developed lies in abandoning the question of reality while turning to explain the experience of the observer with the experience of the observer. This is a fundamental move away from a domain of transcendental ontologies to a domain of constitutive ontologies.

Humberto Maturana (2002)

...it is hard to change the habits of science in order for it to accept that new tools are needed for the transformation of what it means to conduct research on mind and for the training of the next generations.

Francisco Varela (1996)

V této kapitole budeme zkoumat jaké jsou meze metafyzických předpokladů současné vědy a filosofického mainstreamu. Konkrétně budeme zkoumat předpoklad existence objektivní reality a předpoklad existence metafyzického subjektu. Tyto předpoklady budeme analyzovat z hlediska enaktivního přístupu v 1. osobě, tedy z hlediska reflexivní zkušenosti. Jde o téma zajímavé zejména pro filosofy, ale i pro pedagogy, psychoterapeuty a další, kteří se zabývají transformací prožívané zkušenosti.

3.1 Supervenience

*otázkou pro některé vědce a filosofy je
proč ve světě vzniká vědomí
otázkou pro mé vědomí je
proč v něm vzniká svět*

autor

V současné době neustále mohutní výzkumný program na pomezí neurověd a filosofie, který se zabývá výzkumem vědomí, tzv. consciousness studies. Z hlediska enaktivního přístupu můžeme vědomí chápat jako poznávání v 1. osobě (resp. přesněji jako reflexivní zkušenost).

Prvořadým úkolem je vysvětlit interakci mezi poznáváním ve 3. a v 1. osobě, tj. interakci mezi fyzikálním pozorováním poznávacích procesů a reflexivní zkušeností. Jde o tzv. psycho-fyzický nebo také mind-body problem. Společným průnikem „mainstreamových“ filosofických přístupů k vědomí je teorie supervenience. Ta vychází z metafyzického realismu a v podstatě jde o předpoklad, že stavy objektivní tj. fyzikální reality kauzálně určují stavy vědomí, tj. že každému stavu fyzikální reality je možné přiřadit právě jeden stav vědomí (včetně absence vědomí), což nemusí platit naopak (tj. existuje možnost přiřadit jeden stav vědomí více fyzikálním stavům). Vědomí je nějakým způsobem vytvářeno (supervenuje nad) jistými objektivními strukturami (např. mozkiem). Supervenience předpokládá, že vědomí je nějakým způsobem jednotné s fyzikální realitou, avšak uspokojivě nevysvětluje, jakým způsobem je vědomí vytvářeno. Různé pokusy

o vysvětlení se často opírají a koncept emergence. Tzv. slabá emergence má dobrý smysl. Popisuje vcelku triviální skutečnost, že struktura vztahů mezi částmi může díky existenci zpětných vazeb vést k nelineárním vlastnostem celku, (tj. vlastnosti nejsou výsledkem prostého součtu vlastností částí). I když je tento fakt triviální, může vést k velmi netriviálním jevům. Jako příklady se uvádí tekutost vody (nebo spíše jevy vznikající v důsledku povrchového napětí), růst krystalů, organizace mraveniště nebo mnohobuněčný organismus. Všechny tyto jevy jsou součástí zkušenosti pozorování fyzikálního světa. Takzvaná silná emergence se však podobným způsobem snaží ze zkušenosti fyzikálního světa vysvětlit vznik vědomí, tj. vlastně veškerou zkušenost. Tj. snaží se vysvětlit veškerou zkušenost na základě části zkušenosti.

Většina filosofů a vědců zastávajících nějakou variantu superveniencie doufá, že vztah mezi objektivní realitou a vědomím vysvětlí budoucí vývoj vědeckého poznání. Avšak superveniencie je metafyzický koncept (rozvíjí vlastně jen předpoklady metafyzického realismu) nikoli vědecká hypotéza. Není ji možné vyvrátit. K tomu bychom museli ukázat, že nějakému fyzikálnímu stavu přísluší více stavů vědomí. Jakým způsobem bychom ale mohli pozorovat, že došlo u jednoho a toho samého fyzikálního stavu došlo ke změně supervenovaného vědomí? Nemůže to být např. verbální report, pohyb ani změna metabolismu, neboť jde o jistý typ fyzikální pozorování. Existence odlišných stavů vědomí je přístupná pouze v reflexivní zkušenosti nikoliv ve fyzikálním pozorování.

Existují však dobré intuice, při nejmenším co se týká verbálních reportů, které naznačují, jak odlišná může být zkušenost spojená s jedním a tím samým výrokem. Např. astrofyzik a filosof Piet Hut, o jehož myšlenkách se ještě zmíníme, uvádí (Hut, 2006b) příklad radikální transformace zkušenosti, kterou však nelze vyjádřit novými slovy. Například je zásadní rozdíl když výrok „Dnes je opravdu krásně“ prohlásí člověk, který před chvílí o vlasek unikl smrti a člověk, který se jen na chvíli zastaví v kolotoči běžných každodenních činností.

Pokud domyslíme supervenienci do důsledků narazíme na řadu dalších absurdit.

V rámci teorie superveniencie není možné vyloučit, že existuje fyzikální stav, který neumožňuje, aby byl poznán. Tj. existuje takový fyzikální stav, k němuž supervenující vědomé obsahy jsou právě takové, že tento stav nepoznávají nebo nereprezentují.

O supervenienci se většinou uvažuje jen v rámci vědomí jednotlivce. Nic nám ale nebrání uvažovat o supervenovaných stavech mnoha či všech vědomí. V takovém případě se můžeme otázat, zda

můžeme vyloučit takový fyzikální stav, kde supervenované obsahy vědomí budou reprezentovat nějaký jiný stav (budou se mýlit).

Odpověď je jednoznačná. Nemůžeme. Nemůžeme dokonce vyloučit ani to, že k takovému omylu všech supervenovaných vědomí nedochází neustále!

Prvním důvodem je již fakt, že k jednomu supervenovanému stavu může náležet více fyzikálních stavů. Nemáme naprosto žádnou možnost zjistit, zda k jistému stavu vědomí přináležejí jeden či více fyzikálních stavů. Pokud jich je více, pak je v nejlepším případě správně reprezentován právě jen jeden z nich

Pokusme se zjednodušit si situaci a prohlásit, že k jednomu supervenovanému stavu vědomí bude náležet vždy jen jeden fyzikální stav. I tak můžeme dojít k absurdním tvrzením, která není možná ani dokázat ani vyvrátit.

- Tvrdím, že slunce začalo vysílat nový druh záření, který způsobuje to, že nevěříme, že se něco takového stalo a zabraňuje nám pochopit a poznat toto „záření“.
- Tvrdím, že jistá skupina vědců (mimozemšťanů, umělá inteligence apod.) spustila zařízení, které uvedlo většinu lidí na Zemi do stavu, ve kterém žijí ve virtuální realitě, která je nerozeznatelná od skutečnosti. V této virtuální realitě právě teď žijeme. Ona skupina vědců se mezitím stará o naše těla a předešla nebezpečím, která hrozila od nekontrolovaných průmyslových (jaderné elektrárny apod.) a vojenských zařízení.
- Tvrdím, že skupina vědců provedla experiment, při kterém modifikovala vědomí všech lidí na planetě, včetně svého vlastního vědomí, takovým způsobem, že jsou přesvědčení, že se tento experiment nezdařil a zdařit ani nemohl. V takovém případě pak není možno výsledek experimentu objektivně vyhodnotit - buď se povedl a jsme přesvědčení, že se nezdařil, nebo se nepovedl – a pak může být naše přesvědčení jakékoliv.

Soustředme se na poslední tvrzení, evidentně jde o variantu známého sebe-referenčního paradoxu typu „Tato věta není pravdivá“. Tuto větu není možné označit ani za pravdivou ani za nepravdivou. Nebo můžeme prohlásit, že její pravdivostní hodnota se mění ve zpětnovazebném cyklu (pokud je pravdivá pak není pravdivá, pokud není pravdivá, pak je pravdivá atd.). Předpoklad existence objektivní reality, je zároveň předpokladem, že platí zákon vyloučeného třetího, tj. buď nějaký proces v objektivní realitě nastal nebo nenastal, třetí možnost neexistuje. Pokud se však naše tvrzení odvolává mimo zkušenost, nemůžeme zabránit tvrzením u kterých pravdivost a nepravdivost nelze rozhodnout. Pokud bychom přestali lpět na metafyzických

předpokladech mohli bychom jednoduše konstatovat, že takové výroky nemají smysl, respektive není možné určit jejich pravdivostní hodnotu.

Výše zmíněné intuice a absurdní důsledky samozřejmě supervenienci nevyvrací. Metafyzická tvrzení pokud jsou vnitřně logicky konzistentní není možné ani dokázat ani vyvrátit, jsou to předpoklady, které rámuji náš pohled na svět. Kritériem pro přijetí nebo odmítnutí metafyzických předpokladů je, zda nám umožňují smysluplně interpretovat zkušenost. Zejména jde o to, aby se metafyzické předpoklady nestaly implicitní a nereflektované, a nevedli k rigidní a omezené interpretaci zkušenosti.

Tento typ předpokladů se stane nejzřetelnější pokud začneme zkoumat čistě reflexivní zkušenost a zaměříme pozornost na procesy, ze kterých se naše zkušenost skládá.

3.2 Reflexivní zkušenost

Otázka „co poznávám?“ se v konečném důsledku mění v otázku „kdo či co jsem?“ – a provokuje odhalení bytí, které bylo až dosud v anonymitě a nevědomé. V okamžiku přímého, přesného a upřímného položení této otázky – tedy po důkladném sebe-prozkoumání tázajícího se metodickou reflexí – dojde navíc k čemusi neočekávanému: bytí tázajícího se se položením této otázky promění: ten, kdo tuto otázku položil, není již tím, kdo porozuměl jejímu smyslu. Proměnu existence toho, kdo se dotazuje na bytí samo, nazýváme v této práci transformací. Transformace skrze reflexi však není čímsi nahodilým a zmateným, ale má svou přesnou dynamiku, jejíž popis nenalezneme nikde v dějinách západní filosofie.

Michal Šašma (2008)

Protikladem předpokladu existence objektivní reality je idealistický předpoklad existence metafyzického (substanciálního) subjektu, Já, duše apod. Pokud bychom se omezili pouze na tento předpoklad a dovedli ho do důsledků, skončíme bychom buď u subjektivního idealismu či solipsismu, tj. u názoru, že celá naše zkušenost je vytvářena naším Já. Nebo skončíme u objektivního idealismu, tj. názoru, že naše zkušenost je vytvářena naším Já, ale to je nějak odvozeno či participuje na objektivním supersubjektu jako je Bůh, nejvyšší Idea apod.

Avšak existence objektivní reality se s metafyzickým subjektem nemusí vylučovat, naopak od Descartových časů se dobře doplňují a jejich kombinace vede k metafyzickému dualismu. Není účelem tohoto článku podrobně rozebírat toto stanovisko, zájemce o rozbor metafyzických předpokladů objektivního realismu (resp. materialismu a fyzikalismu) a různých idealistických systémů odkazuji na (Burian, 2005b). Objektivní realita i metafyzický subjekt spolu sdílejí to, že jsou nezávislé na zkušenosti. Substanciální subjekt si často představujeme jako pozorovatele, který si zkušenost prohlíží, ale není jí sám ovlivněn (tzv. karteziánské divadlo). Tento pohled i když je kompatibilní s lidovou psychologií, byl z různých stran kritizován (např. Dennet, 1991; Varela et al., 1991). Poznamenejme, že předpoklad neměnných (tj. substantiálních či esenciálních) charakteristik subjektivní zkušenosti se v sofistikované podobě nachází i v Husserlově fenomenologii.

Existence metafyzického subjektu není podporována neurofyziologickým výzkumem – neexistuje jediné ohraničené centrum, které by bylo centrem vědomí. Dokonce i projekt hledání tzv. minimálních neuronálních korelátů vědomí je metodologicky problematický (Noe & Thompson, 2004). Hlavně ale existuje celá řada příkladů, které dokládají, že i když ve zkušenosti pozorujeme řadu pravidelností, tyto pravidelnosti nemusí být absolutně stabilní a je možno je měnit.

Člověk se například může přesytit většiny poznávaných slastí. Zkušenost slastí postupně přestává být slastná. Lidé ovšem spoléhají na to, že toto přesycení zmizí a oni budou opět lační jako předtím. Často se však objeví nutnost zvyšovat dávky, aby bylo dosaženo stejné slasti jako dříve.

Podstatně důležitějším příkladem hluboké proměny zkušenosti je učení se různým dovednostem. Učení je do značné míry reflexivní činnost, takže proměnu zkušenosti, ke které dochází díky učení dovednostem je již možno nazvat transformací. Při takovém učení není nestačí pouze popsat výsledek nebo cestu k němu, ale je nutno provádět praktickou přípravu, při které opakujeme, trénujeme činnosti, které se chceme naučit.

Mezit takové dovednosti patří například hra na hudební nástroj, učení se cizím jazykům, manažerské a sociální dovednosti, různé expertní dovednosti atd. Někdy se mluví o tacitních znalostech, tedy o znalostech, které nelze jednoduše vyjádřit pojmovým popisem.

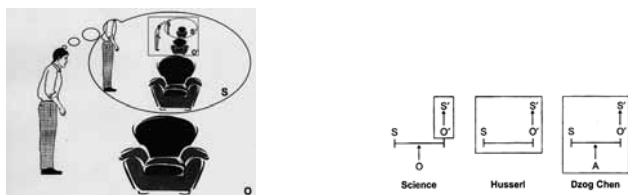
Získávání dovedností obecně je nemožné přesně popsat pomocí nějakého modelu. Ovšemže je možné konstruovat modely zjednodušené, tím ale zhoršujeme naše možnosti naučit se i vyučovat tyto dovednosti. Dokonce i samotné učení se a vyučování je také dovednost.

Uvedu konkrétní ilustrační příklad z vlastní zkušenosti. Nedávno jsem se učil hrát na australský dechový nástroj digeridoo a musel se naučit tzv. cirkulární dýchání, při kterém se zároveň nadechuje nosem a vypouští vzduch z úst. Po dva měsíce jsem metodou pokusů a omylů pomalu zdokonaloval tuto schopnost. Zlepšení se nedělo pomocí pojmového myšlení, ale pomocí nepojmového poznání – vhledu, který spočívá v aktivním a bdělém pozorování prováděné činnosti. Přičemž se zkušenost provádění této činnosti postupně prokresluje, zpřesňuje, získáváme přehled o jemné struktuře průběhu této zkušenosti a tento vhled nám umožňuje zaměřit pozornost na nové složky zkušenosti, kterých jsme si předtím nevšimli. Při učení se hře na hudební nástroj tedy to co poznáváme, tj. naši aktivitu - hraní na nástroj, tímto poznáváním transformujeme. Existuje zde zpětnovazebný cyklus, přestává být odlišitelné, co je poznávání a co poznávané, poznávání hry je samotnou hrou. Tento proces je vlastně reflexivní analogií poznávacích procesů u otevřených agentů, které jsme popisovali v minulé kapitole. Tam se smazával rozdíl mezi poznávajícím agentem a jeho prostředím.

Metafyzické předpoklady jsou v podstatě modely zkušenosti. Každý -ismus, je pojmová (konceptuální konstrukce), kde z jistých konceptuálních konstrukcí - předpokladů či axiomů, jsou logickými pravidly (typicky modus ponens a substituce, ale můžeme mít i předpoklady o tom, jaká logická pravidla budeme používat) odvozovány další konceptuální konstrukce. Různé -ismy se liší právě v předpokladech.

Pokud je zkušenost plně otevřená, tj. plně transformovatelná, tj. je v nějakém smyslu zobecněným případem toho, co jsme v minulé kapitole nazvali otevřený agent, pak můžeme prohlásit, že ji nelze popsat, pochopit, modelovat na základě metafyzických předpokladů.

Otázkou je, zda je možné se bez metafyzických předpokladů obejít. Odpověď se snaží nalézt např. již zmíněný Piet Hut (1999), který možné typy metafyzických předpokladů ilustruje na následujícím schématu.



Obr. 2. Vědecký, fenomenologický a Dzog Chenový přístup ke zkušenosti (Hut, 1999)

Symbol O označuje objektivní realitu. S označuje subjektivní zkušenost. O' označuje zkušenost pozorování (resp. smyslové poznávání), konečně S' je předpokládaná subjektivní zkušenost ostatních bytostí. Schémata v pravé části ukazují různé přístupy ke zkušenosti. Rámečky vyznačují oblasti, kde je z metodologického hlediska možné tyto přístupy uplatňovat.

Věda zkoumá doménu zkušenosti pozorování O' a v rámci psychologie a příbuzných disciplín doménu S'. V doméně O' také můžeme interpretovat poznávací procesy z hlediska enaktivního přístupu ve 3. osobě.

Objektivní realita O nemůže přímo vstoupit do subjektivní zkušenosti, jedná se vždy jen o předpoklad. Přesto se neustále setkáváme s metodologicky zcela chybným ztotožňováním domény O a O' a domény S a S'. Hlavním důvodem je patrně fakt, že když pozorujeme nějaký fyzikální proces, např. křeslo, tak si normálně neuvědomujeme, že pozorované křeslo je výsledkem konstrukce nebo konstituce v naší zkušenosti, ale vnímáme ho „samozřejmě“ jako objektivní danost. To je praktické z hlediska každodenních činností, ale zavádějící z hlediska metodologického.

Husserlovská fenomenologie rozšiřuje sféru výzkumu i na subjektivní zkušenost. Objektivní realitu nepředpokládá, místo toho předpokládá existenci substanciálních struktur subjektivní zkušenosti.

Třetím přístupem, který Hut uvádí je Dzog Chen, což je jedna z tradic vzešlých z buddhistické meditační praxe, přičemž podobné schéma by bylo možné uplatnit i na jiné buddhistické tradice. Dzog Chen nepředpokládá ani objektivní realitu ani existenci substanciálního subjektu, na druhé straně tyto přístupy ani nepopírá, jednoduše je chápe jako možnou součást zkušenosti. Zkušenost však nechápe jako zkušenost subjektivní, ale jako prostou reflexi aktuální zkušenosti. Bez toho aniž bychom předpokládali, že je něco, co je zakoušeno a někdo, kdo zakouší, tj. bez subjekt-objektové duality. Tento stav označuje Hut jako aktualitu A a klade ho do základu Dzog Chenového přístupu. Poznamenejme, že aktualita není novým typem metafyzického předpokladu, ale naopak vyprázdněním všech metafyzických předpokladů.

Jak si ale konkrétně zkušenost aktuality představit? Uvedu zde krátký experiment, ke kterému Piet Hut vyzval účastníky symposia Mind and Reality (Hut, 2006a) a vyzývám laskavě čtenáře/ky, aby si experiment sami vyzkoušeli. Vyberme si nějaký předmět v našem okolí a začněme ho pozorovat, pozorujme ho nejdříve jako zcela samozřejmý objektivně reálný předmět. Potom ten samý předmět pozorujme jako předmět konstruovaný v naší subjektivní zkušenosti. Několikrát tento

přechod od objektivní k subjektivní zkušenosti zopakujme. Po chvíli možná zpozorujeme, že se přitom v naší zkušenosti něco mění – je to jistá složka zkušenosti, která je zodpovědná za vnímání škály mezi objektivitou a subjektivitou. Objektivita a subjektivita jsou pouze složky prožívané zkušenosti, které nemusí být nutně přítomné. Můžeme je s jistou mírou nepřesnosti přirovnat ke zkušenosti hladu a zkušenosti přesycenosti. Oba tyto pocity nějakým způsobem podbarví, ovlivní celou zkušenost. Ale známe i zkušenost kdy necítíme ani hlad ani přesycenost. Pokusme se tedy nyní onu složku zkušenosti zodpovědnou za rozlišování objektivitu a subjektivitu nastavit do neutrálního stavu nebo nejlépe z ní jaksí vystoupit či poodstoupit. Pokud se nám to podaří, získáme vhléd do zkušenosti pozorování předmětu bez metafyzických předpokladů objektivitu nebo subjektivitu¹³.

Plně rozvinout tuto schopnost však není možno dosáhnout hned, ale až po náročném tréninku. Což je do jisté míry „skandální“, protože mnoho filosofů jako by předpokládalo, že schopnost dosáhnout vhlédu do procesů skládajících zkušenosti není nutno nějak zvlášť pěstovat, ale je jaksí automaticky k dispozici. V následující kapitole podrobněji rozebereme metody, které schopnost vhlédu rozvíjejí.

3.3 Transformace zkušenosti

Jak jsem uvedl výše – dokud nám budou metafyzické předpoklady umožňovat smysluplně interpretovat naši zkušenost, není důvod se jich zbavovat. Problém nastane, když počet výjimek, které v rámci metafyzických předpokladů interpretovat nepůjde bude narůstat.

Zatím k tomu u většiny lidí, zdá se, nedochází. U mnoha dovedností jsme vypěstovali postupy, kterými lze člověka vytrénovat kvantitativním opakováním mechanických úkonů. Ovšem způsob jak se utváří skutečný odborník, expert, nedokážeme popsat, přestože si těchto lidí vysoce ceníme. Prostě bereme jako fakt, že někdo „na sobě pracuje“ a „má“ tyto schopnosti. Zachytit podstatu tohoto vývoje je mimo schopnosti pojmového popisu.

Metafyzické předpoklady jsou do nás hluboce vtisknuty vlivem kultury a náboženství. Ovšem různé kultury mají různé předpoklady. Např. euro-americká civilizace, kolébka moderní vědy, bazíruje na

¹³ Tento náznak vhlédu může dát jisté vodítko k řešení mind-body problému, resp. můžeme nahlédnout, že mind-body problém je důsledkem metafyzických předpokladů a při jejich opuštění jednoduše zmizí. Složitější situace je s příbuzným other-minds problémem, tj. problémem existence zkušeností ostatních bytostí. I když i tento problém by měl zmizet při opuštění metafyzických předpokladů, je pro lidi z euro-americké kultury obtížné začít vnímat „ty druhé“ „pouze“ jako součást aktuální zkušenosti a nikoliv jako svébytné zcela nezávislé zkušenosti, tj. zkušenosti jiných substanciálních subjektů. Jedním z řešení je chápat je jako „zrcadla“ která odrážejí aktuální zkušenost a zároveň ji spoluvytvářejí.

předpokladu objektivní reality, zatímco např. hinduismus, má spíše sklony k nějaké verzi objektivního idealismu. Euro-americký individualismus zase vyrůstá z předpokladu existence substanciálního subjektu, zatímco v kolektivistických kulturách Východní Asie, je tento koncept odlišný. V některých tradicích, zejména v některých školách buddhismu, metafyzické předpoklady chybí, resp. je snaha se bez nich obejít. Není náhodou, že právě buddhismus vytvořil praktické techniky rozvíjení nepojmového poznání – vhledu. Hlavní schopností, která dle těchto tradic vhled umožňuje je všímavost (mindfulness, někdy též v češtině překládáno jako bdělá pozornost). Všímavost je rovnovážné spojení velice stabilní a pronikavé pozornosti, kterou je možno zaměřit na libovolnou složku zkušenosti. Rozvinutá všímavost je někdy přirovnávána k lampě, kterou umíme pevně držet a zároveň je dost jasná, abychom si dokázali dobře prohlédnout všechna místa v temné jeskyni (naší zkušenosti). Všímavost je možno rozvinout pouze díky náročnému, většinou roky trvajícimu tréninku. Tento trénink se v sanskrtu nazývá *vipassaná*, což je do češtiny překládáno jako meditace¹⁴ všímavosti a vhledu. Plně rozvinutá všímavost transformuje zkušenost a nepředpokládá žádné limity této transformace, jinými slovy nemáme žádné metafyzické předpoklady o zkušenosti. Tj. nepředpokládáme, že zkušenost, kterou označujeme jako pozorovatel, subjekt, Já apod., je neměnná a nepředpokládáme dokonce, že tato zkušenost je nutná nebo, že její absence nutně musí být patologická.

Současná euro-americká společnost nemá zvláštní tradici metod, rozvíjení vhledu. Ve většině společností, kde tyto tradice existovaly, jsou v současnosti spíše na ústupu. Důvodem snad může být právě pragmatický úspěch vědy a technologií vzniklých v euro-americké kultuře. Na druhé straně si buddhistické metody, získávají pozornost odborníků, kteří se zabývají transformací zkušenosti. Jsou to jednak psychoterapeuti, kteří se snaží o léčebnou transformaci zkušenosti psychicky nemocných lidí, viz např. (Benda 2006). A jednak jsou to pedagogové, kteří se snaží využít všímavosti pro lepší osvojování různých

¹⁴ Existuje velké množství různých typů meditace, jak uvnitř buddhistické tradice, tak mimo ni. Hinduisté praktikují často tzv. transcendentální meditaci, křesťané mají tzv. meditaci srdce apod. V buddhistické tradici jsou meditace rozděleny do dvou základních větví, meditace klidu a meditace vhledu. V různých verzích buddhismu se jejich význam může mírně lišit, avšak základní rozdíl mezi nimi spočívá v tom, že při meditaci klidu se soustředíme na jednu složku zkušenosti, se kterou se snažíme zcela ztotožnit v tzv. bodové koncentraci. Naopak při meditaci vhledu se snažíme dosáhnout stavu, kdy dokážeme soustavně a s velkou přesností sledovat vznikání a zanikání různých složek zkušenosti, přičemž alespoň zpočátku může být jedna z těchto složek hlavní a ostatní vedlejší, sekundární (např. při sledování dechu, si všímám a pozorně zaznamenávám i zvuky, pocity v jiných částech těla, objevující se myšlenky, představy, záměry atd.).¹

ných dovedností od čistě manuálních, až po sebepoznání, sebeovládání, rozvíjení sociálních dovedností apod.

Pro bližší představu uvedme několik příkladů vzhledu získaných pomocí meditace všímavosti a vzhledu. Vycházím z vlastních zkušeností, které se ale z velké části shodují s tím, co je možné se dočíst v příslušné literatuře, např. (Sujiva, 2006).

Nejlépe se vhled projeví u velmi jednoduchých zřetelně sledovatelných tělesných procesů. Např. při pozorném sledování dechu zjistíme po jisté době, že pohyby bránice při dýchání jsou nesmírně složitou soustavou různých typů jednodušších pohybů.

Při sledování bolesti můžeme zase celkem snadno dosáhnout vzhledu, že bolest má dvě složky. Jednak neutrální texturu bolesti – zda je bolest ostrá tupá, jaká má průběh – cuká, pulsuje, nebo je relativně stálá. A jednak negativní složku – tj. nepříjemný pocit, utrpení, chtění, aby bolest již skončila. Tyto dvě složky prožívání je možno v současnosti již do jisté míry identifikovat i neurálně, např. při chronických bolestech byla pacientům s neléčitelným zánětem trojklanného nervu přetřata část mozku v oblasti gyrus cingulum (cingulektomie). Pacienti poté udávali, že sice ví, že cítí bolest, ale ta jim nevadí.

Je možné, po patřičném tréninku, rozvinout i bdělé myšlení – to neznamená descartovské jasné a zřetelné (tj. aktuální a nezkraslené) zaznamenávání výsledků našeho myšlení, ale pozorování samotného myšlenkového procesu, pozorování vznikání zanikání a vzájemné podmíněnosti myšlenkových procesů. Nesnažíme se přitom o zachycení pojmově uchopitelných pravidelných struktur, protože předem existenci takových struktur nepředpokládáme. V kontrastu k tradičnímu pohledu „západní“ filosofie nám bdělé myšlení nám přináší vhled, že myšlenkové procesy jsou zkušenosti, které nejsou v protikladu k smyslovému poznávání či emocím.

Filosofy při čtení předcházejících řádků patrně napadlo, že se jedná o jakýsi typ fenomenologie. V podstatě ano, ale tento problém je velmi složitý. Velmi dobře je rozebrán např. v (Šašma, 2008). Šašma je inspirován zejména Varelou, který Descartovskou tradici meditací a její pokračování u Husserla komentuje následovně:

Domníváme se, že Descartův závěr, že [on, Descartes] je myslící věcí, byl produktem jeho otázky a tato otázka zase byla produktem jeho specifické praxe – praxe od-tělené [„disembodied“], nebdělé reflexe. Husserlovská fenomenologie, ačkoli se chopila prožívání radikálním způsobem, v této tradici pokračovala, protože reflektovala pouze esenciální struktury myšlení. (Varela et al., 1991)

Můžeme získat také vhled do mechanismů, které ovládají naše motivace a chování. Jde zejména o vhled do míry v jaké nás manipulují emoce strachu, vzteku, chtění, zaslepenosti apod. Poznamenejme, že vhled zároveň přináší i možnost rozvíjet aktivní kontrolu nad těmito procesy.

Rozvíjení všímavosti nám pomůže zjistit, že existují další složky zkušenosti, které běžně nereflktujeme. Například to, že se nám něco zdá pravdivé nebo nepravdivé, pravděpodobné nebo nepravděpodobné, že něčemu věříme nebo nevěříme. Toto nejsou hotové danosti, ale opět nějaké typy zkušeností, do kterých je možno získat vhled, zjistit jakou mají strukturu, jakým způsobem jsou konstruovány a případně je transformovat.

Konečně je možno získat vhled i do konstrukce zkušenosti objektivity a subjektivity. Pro buddhistickou tradici je klíčový vhled (nikoli jen informace), že zkušenost mého já je právě jen konstrukce, která není nutná a kterou je možno rozpustit, bez tohoto aniž by výsledek byl patologický stav. Naopak na konkrétních příkladech lidí s vysoce rozvinutou všímavostí je možné ověřit, že výsledkem je stav vyrovnané a zároveň angažované moudrosti.

Avšak představa ztráty substanciálního subjektu vyvolává v lidech euro-amerického kultury strach kořenící patrně už v křesťanských obavách o spásu věčné duše. Tento vhled však není podmínkou pro využívání všímavosti při rozvíjení běžných dovedností. Proto je meditace všímavosti a vzhledu praktikována nejen buddhisty, ale i ateisty či příslušníky dalších náboženství.

V posledních letech se zájem o metody rozvíjení všímavosti prudce šíří, koná se řada konferencí a workshopů (příkladem může být výše zmíněné Mind and Reality symposium), kterých se účastní jak buddhističtí mniši tak filosofové, psychoterapeuti, pedagogové a neurovědci. Důležitou roli v tomto procesu hraje i Dalai Lama, který se již řadu let snaží osobně stimulovat dialog mezi buddhismem a západní vědou. Přesto se stále jedná jen o okrajový proud, který nemá velké šance zasáhnout mainstreamovou společností.

4 Sci-fi místo závěru

Dovedeme si představit situaci, kdy by metafyzické předpoklady začali být citelně nepraktické, a větší část vědců i celé společnosti by byla motivována ke změně paradigmatu?

Domnívám se, že takovou situaci by mohl přinést rozvoj technologií, které umožňují ovlivňovat transformaci zkušenosti. V současnosti je velký zájem věnován tzv. neurokognitivním zdokonalením (neurocognitive enhancement), které mohou zvyšovat schopnosti jako krát-

kodobá paměť, soustředění, ovládaní emocí apod. a jsou dostupné zatím buď ve formě různých farmak nebo tzv. bio-feedbacku. Obě tyto metody jsou velice nespecifické a míra jejich účinku je diskutabilní (zejména u bio-feedbacku). Existuje však reálná perspektiva, že v budoucnu budeme mít k dispozici technologie schopné paralelně stimulovat řadu specifických mozkových center pomocí koordinovaných a zpětnovazebných vzorců stimulačních impulsů. Zpřesnění poznatků o detailním fungování mozku a rozvoj nanotechnologií by pak mohly dokonce umožnit paralelní stimulaci nesmírného množství částí celého mozku (tj. nikoliv např. jen mozkové kůry) na úrovni shluků pouhých desítek či stovek neuronů.

Tyto technologie mají potenciál usnadnit a urychlit učení dovedností a obecně transformaci zkušenosti. Nazvěme proto tyto technologie transformační technologie. Nedomnívám se, že by bylo možné získat nějaké dovednosti instantně, že zapnu aparát a najednou umím hrát na housle. Transformační technologie mohou pouze proces učení zefektivnit, přičemž zůstává prostor pro naše aktivní úsilí. Pokládám za pravděpodobné, že dovedností, kterou se bude třeba naučit, bude samotné řízení těchto technologií, podobně jako je tomu u současných metod bio-feedbacku. Domnívám se, že získávání dovedností je proces, který není možno předem popsat a modelovat. Jak jsem výše zmínil, transformace zkušenosti je zobecněním otevřených agentů, zkušenost je otevřená, nejsou žádné limity transformace. Proces získávání dovedností může vést k předem neodhadnutelným stavům, je nutno neustále reagovat na aktuální zkušenost, je třeba vždy znova zkoušet, která cesta může vést k efektivnímu výsledku. V tomto zkoušení a nalézání bude asi jádro dovednosti v ovládaní těchto technologií.

Rozvoj transformačních technologií by mohl vést k tomu, že význam nepojmového poznávání, vzhledu vedoucího k transformaci zkušenosti, by podstatně vzrostl. Mohl by vzrůst natolik, že by se jasněji projevila nedostatečnost metafyzických předpokladů a poznávání založeného na pojmovém poznávání, formálních popisech a modelech.

Rozvíjení všímavosti, které normálně trvá desítky let a je závislé i na individuálním nadání, by tyto technologie mohly podstatně urychlit a zpřístupnit širokým vrstvám společnosti. To by mohlo vést k rozsáhlým kulturním změnám.

To nás vede k nutnosti zkoumat nebezpečí a etické problémy plynoucí ze samotných transformačních technologií. Je možné je zneužít k manipulaci lidí. Dále hrozí, že se začne rozevírat rozdíl mezi těmi, kteří mají k těmto technologiím přístup a těmi, kteří k nim přístup nemají. Konečně hrozí, že se na těchto technologiích staneme závislými (tak jako na mnoha jiných technologiích). Řešení těchto problémů

nikdo nezná, nové možnosti nesou vždy nová rizika. Něco nám však mohou napovědět schopnosti lidí, kteří dosáhly vysoké míry vhledu.

Pokud nás mají nové technologie nějak zdokonalovat, musíme se zeptat na význam slova dokonalost. Znamená to jen zdokonalení schopnosti uspět v konkurenčním sociálním boji, stát se lepší a ceněnější než druhý, získávat větší podíl na zdrojích, které má společnost k dispozici? Domnívám se, že skutečně efektivní učení dovedností je možné jen při rozvíjení všímavosti a vhledu. Vhled však vede spíše ke zdokonalení vzájemného pochopení, tolerance, spolupráce a vzájemné zodpovědnosti.

Vhled by také mohl přispět k tomu, aby se z těchto technologií nestala berlička, bez které bychom se už nedokázali obejít. Skutečný vhled je správnou metodou možno usnadnit, ale není možno ho žádnou technologií nahradit, vhled zůstává po dosažení trvalý, bez nutnosti ho pomocí technologií udržovat.

Konečně masové rozvinutí všímavosti a schopnosti vhledu by mohlo znamenat protiváhu nesmírně mocných technologií, kterými disponujeme, ale nedokážeme je bezpečně využívat, protože nemáme vhled do motivačních mechanismů, které manipulují naši zkušenost.

Literatura

Benda, J. Meditace, všímavost a nové směry KBT. *Konfrontace – časopis pro psychoterapii*, 17, 3 (65), 2006, str. 132-135.

Burian, Jan. Bdělost a způsoby poznávání. *Acta Universitatis Palackianae Olomucensis- Philosophica 2007*. Olomouc: Univerzita Palackého v Olomouci, 2007. Dostupné na WWW: <http://eldar.cz/honza/articles/bdelost_final.doc>.

Burian, Jan. Samoorganizace a kognice. In: Kelemen, J., Kvasnička V. (eds.) *Kognice a umělý život VI*: 28.5. – 1.6.2006, Třešť. Opava: Slezská univerzita v Opavě, 2006. Dostupné na WWW: <http://eldar.cz/honza/articles/burian_samoorganizae_kognice_final.doc>.

Burian, Jan. Kognice kontra Informace. In: Kelemen, J., Kvasnička V., Pospíchal J. (eds.). *Kognice a umělý život V*: 30.5. – 2.6.2005, Smolenice. Opava: Slezská univerzita v Opavě, 2005. Dostupné na WWW: <http://eldar.cz/honza/articles/burian_kognice_informace.doc>.

Burian, Jan. Stumbling over substance: Why can't we agree about the nature of consciousness?. *Towards science of consciousness 2005*: 17.8.-20.8.2005, Copenhagen. Copenhagen, 2005. Dostupné na WWW: <http://eldar.cz/honza/articles/Burian_TSC2005_abstract_final.doc>.

Burian, Jan. Věda v bezvědomí. In: Kvasnička V., Kelemen J., (eds.). *Kognice a umělý život IV*: 26.05.-30.05.2004, Hradec nad Moravicí. Opava: Slezská univerzita, 2004. Dostupné na WWW: <http://eldar.cz/honza/articles/burian_ve-domi_plusVarela.doc>.

Dennett, D. *Consciousness Explained*. London: Allen Lane, 1991.

Havel, I.M. Přirozené a umělé myšlení jako filosofický problém In: *Umělá inteligence III*. (V. Mařík a kol., editoři), Academia, Praha 2001, s. 17–75.

Dostupné na WWW: <<http://www.cts.cuni.cz/reports/1999/CTS-99-11.htm>>

Hut, P. Exploring Actuality through Experiment and Experience. In *Toward a Science of Consciousness III*, eds. S.R. Hameroff, A.W. Kaszniak, and D.J. Chalmers. Cambridge, MA: M.I.T. Press, 1999, pp. 391-405.

Hut, P. Transcripts of response, *Panel of wisdom. Mind and Reality symposium*. 2006.

Dostupné na WWW: <http://mindandreality.org/pdfs/hut_response_trans.pdf>.

Hut, P. Transcripts of discussion, *Panel of wisdom. Mind and Reality symposium*. 2006.

Dostupné na WWW: <http://mindandreality.org/pdfs/three_discussion_trans.pdf>.

Kauffman, S. *Čtvrtý zákon*. Paseka, Praha a Litomyšl, 2004.

Maturana, H. Autopoiesis, structural coupling and cognition. *Cybernetics & Human Knowing*, Volume 9, Numbers 3-4, 2002, pp. 5-34(30)

Maturana, H., Varela, F. *The Tree of Knowledge: The Biological Roots of Human Understanding*. New Science Library (Shambhala), Boston, 1987.

Noe, A. Thompson, E. Are There Neural Correlates of Consciousness. *Journal of Consciousness Studies*, 11, No. 1, 2004, pp. 3–28

Suber, P. *The Paradox of Self-Amendment: A Study of Law, Logic, Omnipotence, and Change*. Lang Publishing, 1990.

Sujiva. *Základy meditace vhledu*. DharmaGaia. Praha 2006.

Šašma, M. *Postkarteziánská reflexivní filosofie vtělení ve světle buddhistické meditace samatha a vipassaná*. Dizertční práce, dosud nevydáno, předpokládaná obhajoba v r. 2008.

Varela, F., Thompson E., Rosch E.: *The Embodied Mind*. Cambridge. MIT Press 1991.

Varela, F.: Neurophenomenology: A methodological remedy for the „hard problem“. In *Journal of Consciousness Studies*, „Special Issues on the Hard Problems“, J.Shear (Ed.) June 1996. Dostupné na WWW: <http://web.ccr.jussieu.fr/varela/human_consciousness/article01.html>.

Konceptuálne východiská pre model stelesnenej mysle

Igor Farkaš¹

Abstrakt. Zaoberám sa otázkou povahy ľudskej mysle ukotvenej v tele, a vlastnosťami, ktoré by mal potenciálny výpočtový model mať, aby bol čo najviac konzistentný s empirickými znalosťami o fungovaní ľudskeho mozgu-mysle. Navrhovaný konceptuálny rámec pre takýto model sa opiera o konekcionistickú a dynamickú paradigmatu v kognitívnej vede. Z filozofického hľadiska je tento rámec platformou neredukcionistického fyzikalizmu.

1 Úvod

Otázky týkajúce sa mysle a mentálnych procesov (čo bežne nazývame myslením) si ľudia kládli od nepamäti. Každý z nás je presvedčený o existencii svojej vlastnej mysle, lebo ho k tomu vedie vlastná skúsenosť. Každý z nás má interný prístup ku svojim myšlienkam, ktoré mu mysl'ou prebiehajú a k mentálnym stavom, ktoré zažíva. Mysle' je subjektívnym vlastníctvom, ktoré je zvonku neviditeľné. Napriek tomu, a možno práve preto, nezodpovedaných ontologických otázok (akou entitou je mysle') a epistemologických otázok (ako môžeme poznávať túto entitu) je tu stále veľa (Dennett, 1996). Aký druh mysle má dieťa v predverbálnom štádiu vývinu? Existujú aj iné typy mysli než tá ľudská? Ktoré živočichy už možno majú mysle' a ktoré ešte nie? Môže mať mysle' inteligentný robot, skonštruovaný na inom fyzikálnom substráte, alebo vždy ho budeme považovať len za inteligentný stroj? A bude sa to dať vôbec principiálne rozhodnúť?

Zodpovedať na tieto otázky v jednom príspevku je nereálne už len z toho dôvodu, že všeobecne akceptované odpovede nepoznáme. Ja sa pokúsim zúžiť a prezentovať svoj pohľad na podstatu *ľudskej* mysle a predovšetkým jej ukotvenie v materiálnom tele. Dnes už snáď málokto pochybuje o tom, že ľudská mysle' ako nehmotná entita sídli v hmotnom tele, pretože Descartov dualizmus substancii už dnes ťažko možno považovať za vedeckú alternatívu (jemnejšie formy dualizmu, napr. dualizmus vlastností, sú už prijateľnejšie). Na druhej strane, redukcionistické teórie identity mozgu-mysle alebo eliminačného materializmu tiež nie sú podľa mňa tým „pravým orechovým“, lebo úplne eliminujú a zavrhujú napr. fenomenologické vlastnosti mysle. Tu prezentovaný prístup

¹ Fakulta matematiky, fyziky a informatiky, Univerzita Komenského, Mlynská dolina, 84248 Bratislava a na čiastočný úväzok Ústav merania SAV, Dúbravská cesta 9, Bratislava, e-mail: farkas@fmph.uniba.sk

možno zaradiť k platforme *neredukcionistického fyzikalizmu* (Davidson, 1970), lebo ten síce kladie dôraz na hmotný svet, a teda sa snaží vysvetliť kauzálne mechanizmy len v rámci neho (kauzálna uzavretosť), avšak berie do úvahy, že opis mentálnych vlastností (ich epistemologické poznanie) a ich vysvetlenie nie je redukovateľné na opis jazykom prírodných vied. V tejto súvislosti možno niekomu napadne otázka, že načo potom existuje mentálna úroveň (mysel'), keď takýto prístup ju vlastne nepotrebuje, pretože myseľ nemá kauzálny vplyv na mozog, čo sa javí byť v rozpore so subjektívnou skúsenosťou. Rozpor je podľa mňa zdanlivý; argumenty uvádzam v závere príspevku (časť 6).

Keďže vychádzam z predpokladu o kauzálnej uzavretosti fyzikálneho sveta, budem nevyhnutne hovoriť aj o mozgu (resp. celom nervovom systéme), a o procesoch v ňom, pretože tie by mohli zodpovedať za to, čo sa deje v mysli človeka, aj keď „len“ z pozície tretej osoby (3PP).² Pri vedeckom štúdiu mozgu máme k dispozícii experimentálne prístupy skúmajúce správanie človeka, moderné zobrazovacie metódy (činnosti) mozgu, ktoré umožňujú lokalizovať aktivitu mozgu pri mentálnej činnosti, a výpočtové modelovanie, ktoré umožňuje explicitne skúmať neurálne mechanizmy na zvolenej úrovni abstrakcie. Pozícia prvej osoby (1PP), typická skôr pre fenomenológiu (Havel, 2007) je však, myslím si, tiež dôležitá, pretože nie je redukovateľná na opis z 3PP. Vyžaduje si alternatívne prístupy skúmania, ktoré by mali byť spoľahlivejšie ako introspekcia zavedená v 19. storočí (Chalmers, 1999). Niektorí filozofi vidia väčšie prepojenie oboch perspektív. Napríklad Searle (1999), ktorý chápe mentálne stavy ako čisto fyzikálny fenomén, pretože tie sú produktom mozgu, argumentuje, že subjektívnosť ontológie mentálnych stavov mysle nie je prekážkou ich epistemickej objektívnosti poznávania (príklad: napriek tomu, že vnímanie bolesti je číro subjektívne, môžeme mať objektívnu vedu skúmania bolesti).

2 Exkluzívne fundamenty ľudskej mysle

Dôsledkom toho, že naša myseľ neexistuje „v éteri“, ale je ukotvená v materiálnom svete je fakt, že naše mentálne stavy dokážeme skrývať pred vonkajším svetom len čiastočne, pretože naše telo ich *prezrádza*. Niektoré naše vnútorné stavy (hlavne tie, ktoré sú výrazne citovo podfarbené) do istej miery odkrývame (zámerne alebo nechťiac) smerom k iným subjektom, a to svojimi gestami, náznakmi,

² Dovolím si požičať od pána Havla užitočné skratky 3PP (third-person perspective) a 1PP.

či správaním všeobecne.³ Na základe týchto indikátorov každý z nás dokáže čiastočne čítať mentálne stavy iného subjektu, čo potvrdzuje priame prepojenie nášho vnútra a vonkajších prejavov nášho tela. Schopnosti porozumieť vnútru iného subjektu na základe týchto behaviorálnych indícií sa odborne hovorí *teória mysle*, ktorou disponujeme pravdepodobne ako jediný živočíšny druh. Veľký význam pre vysvetlenie jej fungovania bolo objavenie tzv. zrkadliacich neurónov (v premotorickej frontálnej kôre), ktoré sa pravdepodobne podieľajú na neurálnom ukotvení teórie mysle (Rizzolatti a spol., 1996). Tieto neuróny sú aktivované nielen, keď človek sám vykonáva nejakú činnosť, ale aj keď pozoruje iného človeka pri tej činnosti. Vtedy zrkadliace neuróny reagujú na pozorovanú činnosť a umožňujú takto pochopiť správanie iného človeka, vcítiť sa do jeho perspektívy.

Schopnosť človeka „čítať“ myseľ iného človeka úzko súvisí so schopnosťou imitácie, ktorá podľa najnovších interdisciplinárnych poznatkov pravdepodobne nie je nízkoúrovňovou nenáročnou kognitívnou aktivitou, ako sa v minulosti predpokladalo (Hurley & Chater, 2005). Ďalšiu dimenziu pre výmenu myšlienok (obsahov myslí) u človeka predstavuje prirodzený jazyk. Táto komplexná verbálna forma komunikácie predstavuje obrovský evolučný krok dopredu, lebo nám umožňuje ešte viac odhaľovať naše vnútorné stavy a vymieňať si poznatky o čomkoľvek, čo je verbalizovateľné. Je pozoruhodné, že v blízkosti premotorickej frontálnej kôry sa nachádza tzv. Brocova oblasť, ktorá je zase aktívna pri používaní jazyka. Ponúka sa tu hypotéza, že možno práve táto časť frontálnej kôry a jej evolučný vývoj (spojený s expanziou) zohrali významnú úlohu pri vzniku trojice súvisiacich schopností (čo sa týka vzťahu k iným) charakteristických pre ľudskú myseľ – prirodzeného jazyka, schopnosti imitovať a schopnosti porozumieť mentálneho stavu iného človeka. K tomu ešte môžeme pridať ďalšie schopnosti, ktoré nie sú nevyhnutne viazané na vzťah k inému subjektu. Dokážeme abstraktne myslieť, usudzovať, plánovať, ale aj vykonávať zložité motorické pohyby (ručné práce, tanec, hudba, šport), čo nedokáže žiaden iný živočích. Všetky tieto činnosti v sebe skrývajú schopnosť mozgu spracovávať sekvencie dát, čo vystihuje hypotéza o „sekvenčnej kôre“ (Calvin & Ojemann, 1994, kap. 16).

Naše exkluzívne ľudské schopnosti sú produktom mozgu a súčasne prejavmi mysle. Utvrdzujú z nás v presvedčení, že aj napriek tomu, že ja nemám ako zistiť, že aj Vy máte myseľ, predpokladám to na základe podobnosti (či kvalitatívnej rovnakosti v rámci ľudského druhu). Prirodzenou ľudskou snahou je pochopiť podstatu našich men-

³ Samozrejme, nejaký chladnokrvný jedinec, či skvelý herec sa dokáže navonok tváriť indierentne.

tálnych prejavov, a snahou niektorých je aj reprodukovat' ich (alebo aspoň niektoré z nich) v nebiologickom svete. Napríklad, teória mysle sa stala predmetom záujmu aj pri návrhu inteligentných agentov (Wiederermann, 2003). Nato, aby sme pochopili mechanizmy mentálnych procesov, sú veľmi dobrým pomocníkom výpočtové modely. Čím exaktnejší model, tým lepšie. Na rozdiel od menej detailných teoretických modelov typu „box-and-arrow“ sprevádzaných verbálne formulovanou teóriou, sú exaktné výpočtové modely vďaka svojej matematickej explicitnosti jednoznačné, čo sa týka ponúkaného vysvetlenia mechanizmov, a navyše ponúkajú testovateľné predikcie.

3 Výpočtové prístupy k modelovaniu mysle

Naše vlastné mentálne stavy (presvedčenia, ciele, náklady, atď.) a procesy (myslenie) sú to jediné, k čomu máme prístup. Aj keď zrejme drvivá väčšina z nás bude súhlasiť, že ľudská myseľ je stelesnená (resp. vtelená) v hmotnej realite, nie každý bude asi súhlasiť, že pri snahe pochopiť fungovanie mysle musíme nutne hovoriť o mozgu a neurálnych procesoch. Tento pohľad o existencii potenciálne nezávislej, vyššej úrovne ontologickej reality existoval už v antickej filozofii mysle, a v polovici minulého storočia bol výrazne posilnený nástupom počítačov (analogia mysle-mozgu so softvérom-hardvérom počítača), ktoré významne posilnili ovplyvnili názorové spektrum v kontexte odvekého problému vzťahu mozgu a mysle (Farkaš, 2007). Inými slovami, dá sa argumentovať, že napriek tomu, že mozog je sídlom našich mentálnych stavov, na ich opis nemusíme používať neurálne koncepty, ale mentálne koncepty, a ukotvenie modelu mysle je len vecou implementácie. Tento pohľad bol posilnený aj zavedením troch nezávislých úrovní abstrakcie – výpočtovej, algoritmickej a implementačnej⁴ (Marr, 1982), čo bolo motivované práve počítačovou metaforou mysle. Marrov dôraz na prvú a druhú úroveň bol konzistentný s počiatočnými trendmi v umelej inteligencii, kognitívnej psychológii a kognitívnej vedy, ktoré boli postavené na predpoklade, že na implementačnej úrovni abstrakcie nezáleží. Používanie počítačovej metafory mysle je síce možno nápomocné ako veľmi hrubá predstava o fungovaní mysle, no z istého pohľadu môže byť dosť zavádzajúce. Že počítač je počítačom, a teda zariadením realizujúcim výpočty, je zrejmé, no či aj mozog je počítač, také zrejmé nie je. Prečo?

3.1 Robí ľudská myseľ-mozog výpočty?

Výpočet (nejaký algoritmus) možno realizovať rôznymi spôsobmi, počnúc perom a papierom, a končiac implementáciou v počítači. Po-

⁴ Príklad: Triedenie čísel (výpočtový cieľ) možno realizovať rôznymi algoritmami, pričom každý sa dá implementovať v nejakom strojovom kóde (už závislom na hardvéri počítača).

hľad na mentálne procesy pomocou zavedenia formálneho opisu má nesporne svoje výhody, lebo umožňuje navrhovať a študovať mechanizmy, v snahe vysvetliť správanie ľudí (porovnaním s predikciami modelu). Ak sú mentálne procesy výpočtom, potom sa musíme opýtať: akým typom výpočtu, s čím, s akými entitami (reprezentáciami)? Ak ide o výpočty, nemožno takéto výpočty realizovať na inom hardvéri než je mozog? Môže mať napríklad stroj mentálne stavy? Otázka, či mozog-počítač by bol výpočtovým strojom je triviálna, pretože počítač bol navrhnutý ako výpočtový stroj, ktorý komunikuje s prostredím pomocou svojich vstupno-výstupných zariadení, ktoré fungujú ako prevodníky – prevádzajú aktivitu z jedného média do druhého (Dennett, 1996). Spracovanie informácie vnútri počítača je teda „homogénne“ (ide o prenos elektrických aktivít). Okrem toho, v počítači je jednoznačne určená implementačná úroveň, na ktorej výpočty prebiehajú (logické hradlá). V prípade mozgu však implementačná úroveň vôbec nie je jednoznačná, pretože mozog nikto nezostrojil za účelom počítania. Ten sa evolučne vyvinul tak, aby umožnil organizmu efektívne fungovať v dynamickom prostredí (Churchland, Sejnowski, 1992). Na akej úrovni potenciálne prebiehajú výpočty v mozgu? Najpriateľnejšou hypotézou sa javí úroveň neurónov, hoci existujú neurofyziologické poznatky o tom, že nižšia (molekulárna) úroveň fyzikálnej reality ovplyvňuje vyššiu úroveň, t.j. činnosť samotných neurónov (Bell, 1999). Podobne, rôzne (pomalšie) hormonálne a neuromodulátorové systémy ovplyvňujú správanie neurónu (Kaczmarek & Levitan, 1987).⁵ Okrem toho, spracovanie informácie v mozgu má elektricko-chemický charakter⁶ a de facto sa opiera o fyzikálne procesy, ktoré dokážu fungovať v časovej škále optimalizovanej pre existenciu človeka v dynamickom prostredí. Taktiež na rozdiel od počítača, kde môžeme izolovať hardvér od softvéru, a kde hardvér dvoch počítačov toho istého typu je rovnaký, každý mozog je jedinečný, a tak konkrétne mentálne prejavy sú dosahované individuálnymi konfiguráciami fyzikálnych stavov. Na tom, z čoho je skonštruovaný mozog, teda možno záleží.

Napriek nejasnosti, či mozog (ako realizátor mentálnych procesov) možno považovať za počítač, ho budeme v ďalšom texte považovať za výpočtový stroj, čo je dominantným prístupom pri modelovaní. Jedine tak môžeme modelovať mechanizmy kognície či mentálnych procesov v mozgu.

⁵ Mozog nie je len informačný systém, ale aj hormonálny, a teda aj emocionálny systém (LeDoux, 1996).

⁶ Pri šírení informácie v mozgu sa elektrické šírenie aktivity pozdĺž axónu (neurónov) strieda s chemickou difúziou na synapsách (medzi neurónmi).

3.2 Vlastnosti ideálneho modelu mysle

Ak chceme uvažovať nejaký prijateľný model ľudskej mysle, ten musí mať viacero vlastností, ktoré vyplývajú z našich poznatkov. Budem predpokladať, že ideálny model ľudskej mysle by mal mať minimálne tieto vlastnosti:

- A) *robustné správanie* – postupná degradácia fungovanie v prípade šumu či čiastočných porúch systému
- B) *adaptívnosť* – schopnosť učenia, prispôsobovania sa meniacim sa vonkajším podmienkam (prostredia)
- C) *reprezentácia znalostí* – model musí mať vnútorne reprezentované nadobudnuté poznatky, ktoré vie využívať, a ktoré neustále aktualizuje (počas učenia)
- D) *empirická prijateľnosť predikcií modelu* – správanie modelu musí byť v zhode s prejavmi správania človeka. Mám na mysli modelovanie procesov:
 - 1. *nižšej kognície*: vnímanie, rozpoznávanie obrazcov, konanie, senzomotorická koordinácia, a i.
 - 2. *vyššej kognície*: usudzovanie, plánovanie, riešenie problémov a i.
 - 3. *všetkých ostatných procesov*, ktoré nevieme jednoznačne zaradiť, ako napr. používanie jazyka,⁷ priestorová kognícia, imaginácia, pamäť a i.
- E) *vysvetlenie recipročnej kauzality* medzi myslou a mozgom-telom – pretože nás k tomu nabáda vlastná skúsenosť,
- F) *vysvetlenie subjektívnosti* – do akej miery, a či vôbec, vie model vysvetliť, prečo vôbec existuje IPP.

Snaha navrhnúť takýto univerzálny model je asi iluzórna, preto sa budem snažiť len vyčleniť manévrovací priestor a hľadať pre takýto model konceptuálne východiská, konzistentné s čo najväčším balíkom empirických poznatkov. V rámci štartovacej pozície si zosumarizujem poznatky troch hlavných paradigiem v kognitívnej vede, ktoré sa od jej vzniku v polovici minulého storočia postupne vyvinuli. Jednotlivé paradigmy – symbolová, konekcionistická a dynamická, ponúkajú odlišné odpovede na podstatu kognitívnych procesov u človeka.

⁷ Zaradenie jazyka do kategórie D3 je možno pre niekoho trochu máťúce, ale treba si uvedomiť, že schopnosť používať jazyk v sebe zahŕňa procesy na rôznych úrovniach, počnúc od percepcie slov, cez syntax, sémantiku až po diskurz (schopnosť porozumieť jazykovému vstupu s využitím znalosti o svete).

4 Hlavné paradigmy v kognitívnej vede

4.1 Symbolová paradigma

Vznik symbolovej, klasickej paradigmy bol výrazne podnietený vznikom a štúdiom moderných digitálnych počítačov, pretože klasická paradigma konceptualizuje myseľ ako výpočtový stroj, oddelený od prostredia, ktorý manipuluje s internými symbolmi podľa logických pravidiel, podobne ako počítač. Tento pohľad vystihuje hypotéza o fyzikálnom symbolovom systéme, ktorý „disponuje nutnými a postačujúcimi prostriedkami na všeobecné inteligentné konanie“ (Newell & Simon, 1976, s. 87). Pozornosť sa pritom kladie práve na centrálny procesor, ktorý sériovým spôsobom spracováva dáta uložené v pamäti a vykonáva pritom inštrukcie podľa programu (algoritmu) uloženého v inej časti pamäte. Interakcia systému s prostredím je na periférii záujmu a prebieha prostredníctvom vstupno-výstupných podsystémov, ktoré sprostredkujú vstupy do centrálného procesora (vhodnou transformáciou) a výstupy z neho, napr. v duchu teórie modularity (Fodor, 1983). Inými slovami, myslenie a inteligentné správanie (konanie) sú podľa symbolizmu výsledkom operácií prebiehajúcich na symbolovej úrovni. To znamená, že záleží len na type výpočtu, nie na jeho implementácii. Výpočty by mohli byť realizované aj, povedzme, na nejakom univerzálnom tkáčskom stroji, ktorý by prijímal vstupy v spracovateľnej forme a posielal svoje odpovede výstupnému rozhraniu.

Symbolový pohľad je dodnes mnohými odborníkmi považovaný za ten správny najmä v kontexte procesov vyššej kognície ako je usudzovanie, plánovanie, a tiež používanie jazyka. Jeho výhodou je, že poskytuje matematický a logický formalizmus (napr. propozičná či predikátová logika) pre opis, ktorý je transparentný a preto človeku zrozumiteľný. Operácie používané na symboloch sú blízke ľudskému mysleniu (Pstružina, 2007). Pre človeka je prirodzené uvažovať v termínoch podmienok, inferencií a dôsledkov. Navrhnutých výpočtových systémov na báze symbolovej kognitívnej architektúry existuje viacero (pozri napr. prehľad na en.wikipedia.org). Predstavujú v súčasnosti najschodnejšiu cestu ako modelovať myšlienkové pochody človeka sprevádzajúce úlohy vyššej kognície.

Negatívom symbolových systémov je, že celú znalosť (bázu znalostí i pravidiel) do systému musí vkladať dizajnér, znalý domény. Učenie hrá minoritnú úlohu. Taktiež, kde vznikajú problémy, sú úlohy vyžadujúce kognitívne procesy na nižšej úrovni, ktoré človek takisto poľahky zvláda. Príkladom môže byť vizuálne rozpoznanie nejakého objektu, alebo hlasu človeka. Je zrejme, že v týchto prípadoch sa člo-

vek opiera o vizuálne či akustické charakteristiky vstupu, a na základe ich spracovania dospeje k stavu rozpoznania, spojeného s nejakým mentálnym stavom (*áno, je to on!*). Symbolové systémy sa takéto úlohy tiež pokúšali riešiť, a to prepísaním vstupných dát do symbolovej formy a ich následným spracovaním. Takéto snahy sa však stretli s neúspechom, a to nielen kvôli vlastnostiam dát (vysoká rozmernosť, množstvo, variabilnosť, redundancia), ale hlavne kvôli zložitosti navrhnuť veľké množstvo pravidiel, ktoré by zabezpečili aspoň prijateľné fungovanie systému, a ktorých dizajn leží na ramenách programátora. Tu sa priamo ponúka prístup „zdola,“ čo je charakteristickou črtou konekcjonistického prístupu.

4.2 Konekcjonistická paradigma

Konekcjonistický systém je inšpirovaný architektúrou a fungovaním mozgu. Pozostáva z množiny jednoduchých prvkov – umelých neurónov, navzájom komunikujúcich cez váhované spojenia (váhy), ktoré predstavujú dlhodobú pamäť systému. Sila konekcjonistického systému (neurónovej siete) nie je v samotných neurónoch, ale v ich *vzájomnom prepojení a interakcii*, ktorá môže byť excitáčna alebo inhibičná. Paralelné spracovanie a distribuovanosť aktivity predstavujú základný architekturný rozdiel v porovnaní so symbolovým systémom. Taktiež, povaha komunikácie medzi neurónmi má skôr *numeric* než symbolový charakter. Numerickej charakter komunikácie je síce tiež istou abstrakciou od dejov v biologických neurónoch, ale je veľmi vhodným nástrojom na modelovanie kognície, pretože mnoho kognitívnych procesov má kvantitatívne prejavy. Neurónová sieť teda nespracováva symboly, ale numerickej dáta vo vektorovej forme. Operácie nad takýmito vektorovými dátami sú (nelineárnymi) *transformáciami vektorov*. Je zjavné, že takéto operácie nie sú vôbec transparentné a preto sú človeku málo zrozumiteľné.

Z pohľadu symbolovo-subsymbolovej dichotómie môžeme rozlíšiť dva typy konekcjonistických modelov. *Lokalistické* modely sú podobné symbolovým modelom v tom zmysle, že každý umelý neurón reprezentuje (symbolizuje) nejakú entitu. Avšak v *distribuovaných* modeloch sú jednotlivé entity reprezentované množinou neurónov (tzv. populačné kódovanie), čo má niekoľko výhod: umožňuje *lepšie reprezentovať sémantické koncepty* a jemné rozdiely medzi nimi a vnášajú do systému *robustnosť*. Vďaka robustnosti má systém vlastnosť *postupnej degradácie*, pretože i v prípade šumu či lokálnych porúch si zachová istú funkčnosť, ktorá bude závisieť od miery poškodenia. Na druhej strane, problémom modelov neurónových sietí práve kvôli distribuovanosti informácie je, ako súčasne reprezento-

vať viacero konceptov tými istými neurónmi (tzv. problém viazania).⁸ Druhou ťažkosťou s tým súvisiacou je reprezentácia štruktúrovaných entít (dáta so stromovou, či grafovou štruktúrou), typických pre vyššiu kogníciu.

Na druhej strane, neurónová sieť sa dokáže *učiť*, t.j. meniť svoje parametre (váhy spojení), čo znamená, že tie nemusíme explicitne nastaviť (čo sa prakticky ani nedá), len treba sieti poskytnúť vhodný učiaci algoritmus. Podobne ako človek, sieť sa učí na základe tréningových príkladov. Učiaci algoritmus spôsobuje zmenu váh tak, aby sa po naučení dosiahlo požadované správanie siete, čo formálne odpovedá minimalizácii nejakej chybovej funkcie. V závislosti od toho, s akou mierou explicitnosti sa sieti podáva informácia o tom, akú robí chybu, rozlišujeme tri skupiny učiacich algoritmov. Pri *učení s učiteľom* každý neurón dostáva takúto informáciu, a ten mení potom svoje vstupné váhy tak, aby sa jeho chyba znižovala. Pri *učení odmenou a trestom* len sieť ako celok dostáva informáciu o aktuálnej chybe v podobe skalárnej veličiny. Napokon, pri *učení bez učiteľa* sieť nemá k dispozícii nijakú explicitnú informáciu o chybe, preto sa v tejto súvislosti hovorí o samoorganizácii siete.⁹ Existujú teórie, že všetky tri skupiny učiacich algoritmov by mohli byť relevantné pre rôzne časti mozgu, ale pre mozgovú kôru ako sídlo vyšších kognitívnych funkcií je najpriateľnejšou hypotézou samoorganizácia (Doya, 1999).

Učenie siete priamo súvisí s jej schopnosťou *zovšeobecňovať* (generalizovať) svoje znalosti a správne reagovať aj v nových situáciách. Neurónová sieť hľadá štruktúrne vzťahy počas učenia na (tréningových) príkladoch a túto znalosť potom generalizuje pri testovaní. Neurónovú sieť však možno používať aj spôsobom, ktorý bližšie odpovedá interakcii človeka s prostredím, teda keď učenie i testovanie prebiehajú naraz (človek sa učí celý život a pritom permanentne reaguje v nových situáciách). Čo však drvivá väčšina konekcionistických modelov zdieľa so symbolovými systémami je to, že fungujú ako systém izolovaný od prostredia, ktorý spracováva dáta predložené dizajnérom. Sieť nepôsobí aktívne na prostredie, neinteraguje s ním.

4.3 Dynamická paradigma

Niektorí autori prišli s myšlienkou rozšíriť koncept mentálnych stavov (kognície) vo vzťahu k okolitému prostrediu, pretože viacero

⁸ Tento problém sa dá do istej miery zmierniť priblížením sa k lokalistickým reprezentáciám (pomocou tzv. riedkeho kódovania), resp. odstránením v biologicky realistickejších modeloch s impulznými neurónmi pomocou ich časovej synchronizácie.

⁹ Napriek absencii explicitnej chybovej informácie aj samoorganizujúca sa neurónová sieť môže optimalizovať svoje správanie, a to pomocou extrakcie štatistických vzťahov zo vstupných dát (v závislosti od učiaceho algoritmu).

aktivít, ktoré človek vykonáva, sa dá chápať ako výsledok permanentnej interakcie s prostredím (Thelen & Smith, 1994; van Gelder & Port, 1995). Kognitívne (mentálne) procesy teda nie sú chápané izolovane od prostredia, ale v bezprostrednej previazanosti s ním (coupling). Kognitívne procesy s tým súvisiace sú mentálnou aktivitou v čase, modelovateľnou pomocou systému nelineárnych diferenciálnych rovníc. Táto paradigma, postavená na teórii dynamických systémov predpokladá existenciu dvoch kľúčových, vzájomne prepojených typov premenných. Jedným sú tzv. *kolektívne premenné*, ktoré zahŕňajú vzťahy medzi interagujúcimi zložkami dynamického systému, a vysvetľujú správanie systému. Druhým typom sú *riadiace parametre*, ktorých kvantitatívne zmeny môžu spôsobovať kvalitatívne zmeny v správaní systému (tzv. fázové prechody). Príkladom takéhoto systému môže byť pohyb koňa, ktorého rytmus pohybu (vyjadreného periodickými priebehmi kolektívnych premenných) závisí od rýchlosti pohybu (riadiaci parameter). Kôň sa môže pohybovať v každom čase len v jednom rytme, (fázové) prechody medzi nimi nemusia nutne nastávať pri tých istých rýchlostiach, a závisia aj od stavu prostredia (napr. hrbolatá cesta). Iným príkladom môže byť učenie sa chodiť u batoliat, kde taktiež dochádza k bezprostrednej interakcii s najbližším okolím. Dynamický opis je teda postavený na premenných, ktoré treba identifikovať. V prípade bezprostrednej senzomotorickej interakcie s prostredím sa to zrejme dá. Dynamický pohľad na spoznávanie vonkajšieho sveta je úzko spojený so zjednávacím (enactic) prístupom (Varela a spol., 1991), resp. s pohľadom známym ako stelesnená kognícia, v ktorom sa dôraz z vnútorných reprezentácií (vopred daného) vonkajšieho sveta presúva na *vnímanie a jednanie vo svete*, ktorý sa takto spoluvytvára. Tým vzniká prostredníctvom spätnej väzby kruhový proces, pretože zjednávaný svet je prostredím pre zjednávanie, ale zjednávanie je súčasne predpokladom zjednávaného sveta. V rámci tohto pohľadu si každý subjekt (kognitívny systém) zjedná svoj svet sám v procese dynamickej interakcie, a ten bude jedinečný.

Bressler a Kelso (2001) argumentujú, že dynamický pohľad je relevantný aj pre charakterizáciu samotných mentálnych procesov (vnútri kognitívneho systému), hoci identifikácia kolektívnych premenných a riadiacich parametrov v tomto prípade je zatiaľ ťažkým orieškom. Podľa teórie dynamiky kortikálnej koordinácie mozog pracuje v metastabilnom dynamickom režime, prechádzajúc permanentne z jedného stabilného režimu do druhého (čo odpovedá rôznym mentálnym stavom).

	Trvanie	Úroveň opisu	Reprezentácie
Symbolová	od 1970	symbolová	lokalistické
Konekcionistická	od 1985	subsymbologová	distribúované
Dynamická	od 1990	subsymbologová	žiadne?

Tabuľka 1. Hlavné paradigmy kognitívnej vedy.

4.4 Podobnosti a rozdiely medzi paradigmami

Tri spomínané paradigmy (Tab. 1) ponúkajú rôzne odpovede na otázky spojené v modelovaní mentálnych procesov. Symbolová a konekcionistická paradigma stoja viac-menej ako komplementárne prístupy, no niektorí odborníci sa snažili poukázať na výpočtovú ekvivalenciu medzi nimi. Z výpočtového hľadiska považujeme Turingov stroj za univerzálny výpočtový stroj, a teda symbolový model by mal schopný realizovať akýkoľvek algoritmus. V našom kontexte však otázkou je, či všetky mentálne procesy sú algoritimizovateľné. Aj keby boli, problémom ostáva, ako vhodný algoritmus nadizajnovať, ak to nejde ručne,¹⁰ ako napríklad v doménach s vysokou dimenziou dát (napr. spracovanie obrazovej informácie). Hoci dopredné neurónové siete dokážu aproximovať ľubovoľnú spojitú funkciu (Hornik a spol., 1989) a rekurentné neurónové siete sú výpočtovo ekvivalentné univerzálnemu Turingovmu stroju (Siegelmann & Sontag, 1991), tieto existenčné dôkazy neriešia otázku ako také siete skonštruovať (problém nastavenia váh siete). Symbolisti sa snažia nastaviť parametre svojich modelov ručne, v konekcionizme to prakticky možné nie je.

Silné a slabé stránky oboch prístupov sa dajú dobre ilustrovať v kontexte modelovania jazykovej schopnosti človeka, ktorá sama osebe obsahuje podprocesy na rôznych úrovniach kognície (od fonológie po pragmatiku). Je pozoruhodné, že jazykový prejav človeka sa ukazuje vo väčšom súlade s prednosťami a ťažkosťami konekcionistických modelov než so symbolovými systémami (Rohde & Plaut, 2003; Farkaš, 2005). Napríklad, človek poľahky chápe jemné sémantické rozdiely v slovách či na úrovni viet, no má problémy so spracovaním zložitých rekurzívnych viet. V symbolovom systéme to býva typicky naopak. Psychologické dáta sú v pomerne dobrom kvalitatívnom a často i kvantitatívnom súlade s predikciami konekcionistických modelov (napr. Rohde, 2002; Mayberry, 2003; O'Reilly & Munakata, 2000). Problémom ostáva efektívne učenie týchto zložitých modelov, ktoré bráni ich zväčšeniu (slovníka a komplexnosti použitej gramatiky).

¹⁰ do istej miery sa dá oprieť o prírodu inšpirované prístupy, napr. genetické programovanie, pri ktorom sa časti kódu (programu) nechajú evolučne vyvíjať

Dynamický prístup je pomerne konzistentný s konekcionizmom. Obe paradigmy predstavujú príklady *kolektívnych systémov* pozostávajúcich z množstva interagujúcich prvkov, čoho výsledkom sú *emergentné javy* sprevádzajúce dynamiku systému. V čom sa obe paradigmy líšia je (1) dimenzionalita systému, ktorá býva v prípade konekcionistického systému omnoho vyššia, a (2) fakt, že len v konekcionistických modeloch boli navrhnuté rôzne mechanizmy učenia (adaptácie), a teda existujú dve, vzájomne sa ovplyvňujúce dynamiky, operujúce na rôznych časových škálach – pomalá (adaptačná) a rýchla (aktivačná). V dynamických modeloch sú síce významné fázoové prechody, ale ako k nim dochádza, resp. ako zabezpečiť adaptáciu je už zložitejšie. Dôraz ostáva na aktívnej dynamike. Okrem toho, neurónová sieť sa principiálne dá navrhnuť a použiť spôsobom, aby interagovala s prostredím (či už simulovanom, alebo reálnom, ak jej implementovaná v agentovi), čím by si mohla svoju reprezentáciu prostredia zjednávať.

Čím sa podľa niektorých jeho zástancov (Thelen & Smith, 1994) dynamický prístup odlišuje od ostatných dvoch paradigiem je jeho ne-výpočtový charakter a to, že postuluje neexistenciu (mentálnych) reprezentácií (tzv. radikálna téza). Je to zdôvodňované tým, že dynamický systém vyvíjajúci sa v čase sa nedá implementovať pomocou diskretných výpočtov, a preto nemôže podporovať ani žiadne vnútorné reprezentácie vonkajšieho sveta. Tento pohľad je možno odôvodniteľný v kontexte modelovania senzomotorickej interakcie človeka s prostredím, zúženej na koordináciu pohybov (kde asi naozaj nepotrebujeme vnútorné reprezentácie). Ťažko ho však podľa mňa prijať v širšom kontexte kognitívnych procesov človeka, ktoré si vyžadujú existenciu mentálnych reprezentácií, či už vyvolateľných (cued) vonkajším podnetom, alebo oddelených (detached), t.j. vyvolateľných vnútorne (Gärdenfors, 1996). Aproximácia spojitých dynamických systémov pomocou modelov s diskretným časom je schodnou cestou, ktorá napr. celkom dobre funguje v biologicky prijateľnejších modeloch neurónových sietí s impulznými neurónmi (Izhikevich, 2004). Preto sa myslím, že reprezentáčno-výpočtový prístup je kompatibilný s „mäkšou“ verziou dynamickej paradigmy, resp. s pohľadom na mozog ako na samoorganizujúci sa systém.¹¹ Podobnú optiku vidieť aj u menej radikálnych zástancov dynamickej paradigmy (Port & van Gelder, 1995).

¹¹ „pattern-forming, self-organizing system“ (Kelso, 1995, s. 26)

4.5 Hypotéza o symbolových operáciách pri pohľade „zdola“

Domnievam sa, že medzi symbolovými a subsymbolovými systémami existuje istá asymetria. Otázka vysvetlenia procesov nižšej kognície je prakticky neschodná pre symbolové systémy ako som argumentoval v závere časti 4.1. Z opačného pohľadu, vysvetlenie mechanizmov sprevádzajúcich vyššiu kogníciu pomocou nejakého adaptívneho kolektívneho systému je možno len otázkou času kvôli ťažkostiam tohto prístupu. Tu vzniká otázka, ktorú Clark (2001, s. 135) označil termínom *kognitívna inkrementálnosť*: „Aký je vzťah medzi stratégiami používanými pri základných problémoch vnímania a konania, a tými, ktoré používame na riešenie abstraktnejších problémov na vyššej úrovni?“ Inými slovami, je medzi nimi kvalitatívny rozdiel (z hľadiska používaných mechanizmov), alebo sú ukotvené v mozgu rovnakým spôsobom na nejakej konekcionisticky relevantnej úrovni abstrakcie? Mohol počas evolúcie nastať nejaký kvalitatívny skok vo vývoji mozgu človeka, ktorý by súvisel s našimi schopnosťami vyššej kognície? Je známe, že evolúcia využíva staré, osvedčené mechanizmy aj pri návrhu nových systémov. Napríklad, biologické mechanizmy plasticity a učenia sú evolučne staré (človek ich má podobné, a v niektorých aspektoch rovnaké, ako nižšie živočích), a v rámci ľudského mozgu vyzerajú byť univerzálne, či už hovoríme o vizuálnej kôre alebo frontálnej kôre (Jedlička a spol., 2002).¹²

Dokážeme ponúknuť hypotézu, ako pomocou kolektívneho systému vysvetliť schopnosť človeka spracovávať aj symbolové dáta? Vo svete okolo nás existujú dáta subsymbolového (numerickeho) charakteru (obrazy, zvuky, atď.) i symbolového charakteru (matematické vzorce, schémy a pod.). Človek nepochybne disponuje schopnosťou spracovávať aj symbolové dáta (aj keď ich de facto prijíma cez sensorické vstupy v subsymbolovej forme). Je však pozoruhodné, že pri aplikácií rôznych transformácií na symboloch (napr. pri matematických úpravách) si človek pomáha *externou pamäťou*, napr. zapisovaním medzivýsledkov na papier pri násobení čísel. Je to zrejme preto, aby bol schopný mentálne realizovať každý krok – symbolovú operáciu ako vektorovú transformáciu (distribúovanej reprezentácie, t.j. vzorca aktivít), lebo takéto jednokrokové transformácie podporuje ľudská kognitívna architektúra.¹³ Druhým dôvodom je naša obmedzená kapacita pracovnej pamäti. Povedané slovami Hutchinsa (1995),

¹² Na konekcionistickej úrovni abstrakcie ide o tzv. hebbovský princíp, na úrovni impulzných neurónov je v súčasnosti najpriateľnejšou teóriou STDP (spike-timing dependent plasticity).

¹³ Teraz mám na mysli mozog bežného človeka, nie génia (ktorý by bol kontra príkladom). No bolo by zaujímavé zistiť, ako robia géniovia výpočty v hlave. Verím, že tiež využívajú heuristiky, obrazy, a pod.

„výpočtová sila ľudského prístupu nezávisí primárne na našom vnútornom informačnom zariadení (mozgu), ale na spolupráci s externou súčasťou technológie, ktorá nám umožňuje realizovať úlohy, v ktorých sme dobrí: rozpoznávanie vzorov, modelovanie jednoduchej dynamiky a manipulácia s objektmi v priestore“. Ak by bola počítačová metafora pravdivá, prečo potom nevieme robiť celé sekvencie operácií na symboloch v hlave?

Podobne je to s usudzovaním. Ukazuje sa, že kognitívna logika je dosť odlišná od matematickej logiky, a že človek sa často opiera o rôzne „mentálne skratky“ v snahe nájsť prijateľné riešenie v krátkom čase (Kvasnička, 2006). Inými slovami, ľudské konanie má často ohraničenú racionalitu. Ľudia taktiež pri premýšľaní obvykle nepoužívajú dlhé odvodenia. Často hneď „vidia“ závery, ktoré prijímajú, čomu sa hovorí „skákanie k záverom“. Až vtedy, keď pri skákaní narazia na nejaké problémy, prechádzajú k zdĺhavejším postupom odvodzovania, argumentácie (Šefránek, 2007). Možno teda pri skákaní využívajú nejaké naučené efektívne schémy, či heuristiky, ktoré sú mentálne implementovateľné (v podobe vektorovej transformácie). Takisto, Johnson-Laird (1983) tvrdí, že naša logická kompetencia nie je len vecou formálnych pravidiel, ale skôr mentálnych modelov, pri ktorých si pomáhame mentálnou vizualizáciou.

Asi zrejme netreba zdôrazňovať, že možnosť externalizácie symbolovej informácie má obrovský význam pre človeka (knihy, web, a podobne). Symbolové systémy majú zelenú pri návrhu inteligentných systémov spracovania informácie, aj keď kognitívne architektúry takýchto systémov budú (pravdepodobne) diametrálne odlišné od ľudskej mysle, ktorou sa v tomto príspevku zaoberáme.

4.6 Kauzalita vo svetle paradigiem

Ťažším orieškom je vysvetlenie recipročnej kauzality medzi myslou a telom (podmienka E), ktorú možno vidieť v kontexte starého filozofického problému vzťahu myseľ-telo (Havel, 2001). Tomuto problému sa venovalo veľa priestoru vo filozofii mysle. Ako môže nejaký formálny model vysvetliť recipročnú kauzalitu? Ja som k tejto otázke zaujal stanovisko v predchádzajúcich príspevkoch (Farkaš, 2006, 2007) a tu ho mierne rozšírim a dám do širšieho kontextu.

V systémoch ktoré sú opisateľné na viacerých úrovniach abstrakcie je možné uvažovať kauzálne vzťahy v dvoch dimenziách:

(a) *na jednotlivých úrovniach* – t.j. v kauzálnych doménach (Havel, 2001), v rámci ktorých môžeme hľadať príčinné vzťahy medzi javmi. Kauzálne chápaný vzťah v jednej doméne (vidím hada, preto znehybním), môže byť vysvetlený sprievodne v inej doméne (neuróny

vo vizuálnej kôre registrujú hada, čo spôsobí vhodnú konfiguráciu aktivít neurónov v motorickej kôre, ktoré blokujú svalové pohyby). Toto môžeme nazvať *kauzálnym paralelizmom* (Havel, 2001, s. 46).

(b) *medzi úrovňami* – napr. v snahe vysvetliť vplyv tela na myseľ a opačne. Napríklad, už dávno sa vedelo, že poškodenia mozgu vedú k mentálnym poruchám. Taktiež, dlhodobé negatívne pôsobenie prostredia na jedinca (na jeho psychiku) môže spôsobiť degeneratívne zmeny v jeho mozgu/tele.

Oba typy kauzality sú relevantné pre kolektívne systémy, kde vzniká priestor pre modelovanie kauzality na úrovni elementov systému (jednosmerné interakcie), ale taktiež „zdola nahor“ (emergenčia) a „zhora nadol“ (podriadenie elementu systémom).¹⁴

Vysvetlenie kauzality v kontexte symbolovej a konekcionistickej paradigmy má rozdielne charakteristiky. Symbolová paradigma, blízka ľudovej psychológii (psychológii tzv. zdravého rozumu) predpokladá, že „funkčne diskrétné, sémanticky interpretovateľné stavy hrajú kauzálnu úlohu vo vytváraní iných propozičných postojov a celkovo v generovaní správania“ (Ramsey a spol., 1990, s. 220). Napríklad, vplyvná Fodorova (1987) reprezentačná teória mysle predpokladá, že (1) propozičné postoje sú výpočtovo dané do vzťahu s internými reprezentáciami, a že (2) mentálne procesy sú kauzálnymi procesmi, ktoré implikujú prechody medzi internými reprezentáciami (s. 16-20). Aj ľudová psychológia nás nabáda k presvedčeniu, že mentálne stavy majú kauzálnu silu (Fodor, 1987, s. 12).¹⁵ Ľudia sa predsa o ňu každodenne opierajú pri riešení najrozličnejších problémov.

Z redukcionistického pohľadu sú však takéto mentálne kauzálne vzťahy len *užitočnou metaforou*, pretože ide o pseudokauzalitu (Churchland, 1989). Skutočné kauzálne vzťahy sú „roztrúsene“ (scattered) ukotvené v neurálnom substráte (kolektívnom systéme). Preto ľudová psychológia „by mala byť eliminovaná, lebo je zavádzajúca a chimérická“ (s. 127). Menej kritický postoj prezentuje Dennett (1987), ktorý nezatraca ľudovú psychológiu v ľudskej kultúre, pretože verí, že v niektorých prípadoch bola a je pre vedu užitočná (a sa ukáže ako konzistentná s vedeckými teóriami).

Konekcionistické modely sú zase kvôli svojej distribuovanosti spájané s javom, ktorý Ramsey a spol. (1990) nazvali *totálny kauzálny holizmus* (s. 181). Ten podľa kritikov nie je kompatibilný s predstavou o diskretných príčinných vzťahoch typických pre symbolové systémy.

¹⁴ Pri radikálnej verzii dynamickej paradigmy sa recipročná kauzalita rozplýva v podobe kontinuálnej interakcie medzi mozgom, telom a prostredím v reálnom čase.

¹⁵ Ako inak, než že sme smädní, by sme si vedeli vysvetliť, že máme intenciu ísť do chladničky pre pivo?

Keďže v neurónovej sieti sú reprezentácie konceptov distribuované, a teda každý neurón sa podieľa na reprezentácii viacerých konceptov, je ťažké identifikovať, aké kauzálne vplyvy majú jednotlivé spojenia v sieti. Obhájenie konekcionistického vysvetlenia kauzálnych vzťahov môže vyzerat' nasledovne: V rámci jednej úrovne (umelých neurónov) sa môžeme opierať o fakt, že aj v distribuovanej sieti existujú štruktúrované závislosti, ktoré možno odhaliť pomocou bežných techník ako napr. zhluková analýza.¹⁶ Štruktúra týchto závislostí vzniknutejších v priebehu adaptácie siete je odrazom štatistických závislostí medzi jednotlivými elementmi distribuovaných reprezentácií.¹⁷ Pri vysvetlení kauzality medzi úrovňami v rámci kolektívneho systému si zase pomôžeme konceptom emergencie a podriadenia („zotročenia“, enslaving) ako sprievodných javov samoorganizácie systému. O medziúrovňovej kauzalite má však v tu prezentovanom ponímaní zmysel hovoriť len v kontexte dvoch fyzikálnych úrovni a nie vo vzťahu mozog-mysel' (časť 6).

4.7 Systematickosť mysle

Fenómén systematickosti ľudskej mysle sa taktiež priamo dotýka spomínaných paradigiem, a najmä pre konekcionizmus predstavuje stále veľkú výzvu (Aizawa, 2003). Možno ho najlepšie vysvetliť pomocou analógie so schopnosťou človeka systematicky rozumieť jazyku.¹⁸ Ak niekto vie povedať, že *Janko ľúbi Aničku*, bez problémov vie vytvoriť aj vetu *Anička ľúbi Janka*, pretože vie, že jazyk sa skladá z častí a vie, ako ich skladať pri tvorbe viet. Systematickosť v jazyku teda implikuje *produktívnosť* (schopnosť generovať teoreticky neobmedzené množstvo viet) a *kompozičnosť* (existencia vnútornej štruktúry viet).¹⁹ V symbolových modeloch je problém systematickosti automaticky vyriešený, pretože reprezentácie a syntaktické operácie nad nimi (podľa pravidiel) priamo umožňujú produktívnosť symbolového systému. V konekcionistických systémoch takéto vlastnosti na prvý pohľad vôbec nie sú zrejmé, a podľa niektorých kritikov ani principiálne možné. Fodor a Pylyshyn (1988) vo svojom vplyvnom critic-

¹⁶ Takéto techniky sa bežne používajú od 90. rokov, napr. pri analýze vzniknutých interných reprezentácií na tzv. skrytej vrstve rekurentnej neurónovej siete.

¹⁷ Ako vhodnú analógiu pre pochopenie distribuovanej kauzality možno uviesť tento príklad: „Globálna depresia spôsobila nárast nezamestnanosti v štáte Ohio“ (Clark, 2001, s. 54). Je to užitočná metafora pre pochopenie vzťahov medzi fenoménmi na úrovni štátu. Skutočné kauzálne účinky sú distribuované, ukotvené v počínaní jednotlivých ľudí, na ktorých dopadajú dôsledky hospodárskej depresie. Podobne by to mohlo vyzerat' v mozgu.

¹⁸ Podľa niektorých psychológov (Fodor, Pinker) je vnútorným jazykom mysle „mentálna“, do/z ktorej sa transformujú verbalizované myšlienky.

¹⁹ Kompozičnosť predpokladá možnosť syntakticky preusporiadať konštituenty reprezentácie.

kom článku sformulovali svoj argument zhruba nasledovne: „Ľudská myseľ je systematická, a preto jej interné reprezentácie sú štruktúrované. Keďže v konekcionistických systémoch neexistujú štruktúrované reprezentácie, tie (systémy) nie sú dobrými modelmi ľudskej mysle.“ Odozva na túto kritiku prišla v podobe niekoľkých modelov, v ktorých autori poukazovali na existenciu systematickosti distribuovaných reprezentácií v rôznych (rekurzívnych) modeloch neurónových sietí (napr. pozri prehľad vo Farkaš, 2005, časť 3.3). Pozoruhodným poznatkom je to, že distribuované konekcionistické reprezentácie majú tzv. *funkčnú kompozičnosť* (van Gelder, 1990) bez nutnosti *konkateatívnej kompozičnosti*, typickej pre symbolové systémy, ktorá vzniká zretazením reprezentácií jednotlivých elementov (konštituentov). Podstatné je, že funkčná kompozičnosť umožňuje realizovať štruktúrne závislé transformácie bez nutnosti extrakcie elementov (z vektorov) a ich preusporiadania (napr. zámena subjektu a objektu).

4.8 Propozície či mentálne obrazy?

Systematickosť sa priamo týka aj jazyka mysle (mentálčiny). Ak niečo také existuje (ontologická otázka), otázkou je, akú to má povahu – či jazykovú (propozičnú) alebo obrazovú. Propozičná mentálčina (Pylyshyn, 1984) má blízko k formálnym jazykom opisu a môžeme pomocou nej vyjadriť takmer čokoľvek. Propozície²⁰ sa dajú vnárat', kombinovať, môžeme zaviesť ľubovoľné typy propozícií (s rôznymi hlavami a typmi argumentov). Napríklad to, že *Janko má rád Marienku, ktorá venčí psa* by sme vyjadrili ako (*Love John (Is Mary (Walk Mary Dog))*). Propozície majú symbolový charakter, umožňujú opísať aj abstraktné entity a v podstate ľubovoľné vzťahy medzi nimi. Zisťovanie pravdivosti propozície (v nejakom mentálnom svete) je záležitosťou interpretácie symbolov a ich vzťahov. Keďže symboly sú amodálne (v arbitrárnom vzťahu s vonkajším svetom), vystupuje tu problém ukotvenia symbolov (Harnad, 1990). Ten sa snažili autori riešiť rôznymi spôsobmi (pozri prehľad vo Vavrečka, 2006).

Na druhej strane, mentálne obrazy opisujú entity sveta (objekty, udalosti) pomocou schém, či (hlavne) vizuálnych obrazov, ktoré zdieľajú charakteristiky s perceptuálnymi systémami (Barsalou, 1999). Keďže sú spriahnuté s percepciou, majú analógový, a teda subsymbolový charakter (Kosslyn, 1994). S mentálnymi obrazmi je konzistentná teória o tzv. situačných modeloch (Zwaan & Radvansky, 1998), ktoré si človek vytvára v mysli pri spracovaní symbolo-

²⁰ napr. vo forme usporiadaných n-tíc (*hlava arg1 arg2 ...*), kde *hlava* označujúca typ propozície má niekoľko argumentov (minimálne jeden), ktorých typ závisí od hlavy.

vého vstupu (jazyka). Inými slovami, aby človek pochopil nejakú vetu, skonštruuje si v mysli obraz situácie a keď tá je konzistentná, človek pochopil význam. V súčasnosti v literatúre neexistuje konsenzus o tom, ktorý typ reprezentácie – propozičný či obrazový, je kognitívne relevantný, pretože sa zdá, že existujú empirické poznatky v prospech oboch (Sternberg, 2003). Neexistencia konsenzu pramení aj z toho, že konkrétne empirické poznatky môžu byť interpretovateľné z oboch pohľadov. Z nášho pohľadu majú mentálne obrazy oveľa bližšie k distribuovaným reprezentáciám v rámci kolektívneho systému než propozície.

5 Kritériá pre vedomé stavy z pohľadu neurovedy

Pri formálnom prístupe k uchopeniu mysle nemáme principiálne možnosť vysvetliť aspekt subjektívnosti (1PP) mentálnych stavov (podmienka F). Máme však možnosť skúmať, aké kritériá musia spĺňať mentálne stavy z pohľadu 3PP. Ak máme na mysli neurálne procesy, hovoríme o tzv. neurálnych korelátoch mentálnych stavov (Metzinger, 2000). Ak pohľad zúžime na výpočtové aspekty, hovoríme o výpočtových korelátoch (Cleeremans, 2005). Zatiaľ čo neurálne koreláty sú viazané na substrát, v ktorom sú ukotvené, výpočtové koreláty (ak sa ukáže, že existujú) potenciálne umožňujú realizovateľnosť mentálnych procesov u človeka aj na inom substráte (funkcionalistický pohľad). A možno má zmysel skúmať oba typy korelátov a vzťah medzi nimi (Cleeremans, 2005). Mentálne stavy z pohľadu mysle ako kolektívneho systému (3PP) chápeme ako distribuované vzorce aktivít (na nejakej úrovni abstrakcie). Niektoré z týchto vzorcov súvisia²¹ s *vedomím* (čo je evidované odpovedajúcim stavom v 1PP), niektoré nie. V mozgu prebieha stále množstvo procesov, ku ktorým nemáme prístup, sú nevedomé (Dehaene & Naccache, 2001). Vedomie teda chápeme ako špecifickú podtriedu možných mentálnych stavov, ktorej charakteristiky by sme radi odhalili. Vedomé mentálne stavy však nie sú nejakou homogénnou podtriedou, o čom nás presvedčajú každodenné skúsenosti. V odbornej literatúre (napr. Edelman, 1989, pozri aj Havel, 2001, časť 1.8) sa dočítame o typickom delení stavov vedomia na (1) primárne vedomie (senzorické), ktoré vzniká pri vnímaní nejakého pozorovaného objektu, (2) reflexívne (sekundárne), vďaka ktorému dokážem reflektovať vnímané podnety vo svojom „ja“ (som to ja, kto vníma, autobiografia), a (3) vyššie (rozšírené) vedomie, ktoré mi umožňuje zahrnúť i časovú dimenziu do vnútorného modelu sveta i seba (Edel-

²¹ Zámerne nehovorím, že spôsobia vedomý stav (čo by evokovalo kauzalitu), nakoľko to chápeme ako jednu entitu, opisovanú z dvoch perspektív.

man, 1989).²² Je možné, že reflexívne vedomie je vyhradené len človeku a niektorým cicavcom (napr. ľudooppy sa v zrkadle spoznajú), a vyššie vedomie je výhradne vlastnosťou ľudskej mysle (Král & Hulín, 2001). Možno schopnosť mysle človeka reflektovať seba samého v časovom kontexte súvisí s jeho jazykovou schopnosťou (Beňušková, 2005).

Pri hľadaní výpočtových korelátov vedomia sa operuje najmä termínmi reprezentácia a (dynamický) proces. Atkinson a spol. (2000) prezentovali taxonómiu výpočtových teórií vedomia práve podľa týchto dvoch dimenzií. Jednou bola tá, či vedomie skôr závisí od špecifických procesov alebo od (neurálnych) reprezentácií so špecifickými vlastnosťami. Druhou dimenziou bola tá, či vedomie je lokalizovateľné v mozgu, alebo či jeho koreláty sú široko distribuované. V rámci tejto taxonómie sa ohľadne prvej dimenzie skôr prikláňam k názoru, že vedomie je špecifickým procesom, avšak ten sa podľa mňa nedá oddeliť od reprezentácií, na ktorých proces operuje (na rozdiel od počítača). Toto prepojenie sa javí ako konzistentné s požiadavkou ukotvenia neurálnych reprezentácií (Ziemke, 2003). Z pohľadu druhej dimenzie si myslím, že vedomie vzniká distribuovanou aktivitou v mozgu,²³ aj keď zrejme hľadanie participujúcich mozgových oblastí bude možné zúžiť o niektoré oblasti, čomu napovedajú poznatky z neurovedy, ako spomeniem nižšie.

5.1 Neurálne koreláty mentálnych stavov vedomia

Bolo identifikovaných niekoľko kritérií vedomých stavov (na úrovni EEG signálu), ktoré sa dajú aj experimentálne skúmať (Seth a spol., 2005). Niektoré z týchto kritérií možno uplatniť aj u cicavcov, ktoré majú niektoré neurofyziologické predpoklady pre vedomie (Král & Hulín, 2001). Môžeme sa oprieť o tri základné fyziologické fakty o mozgu:

(1) Už pomerne dávno bolo pozorované, že vedomé (bdelé) stavy sú sprevádzané nepravidelným priebehom EEG s malými amplitúdami vo frekvenčnom rozsahu 20-70 Hz, zatiaľ čo počas nevedomých stavov (hlboký spánok, vegetatívne stavy, a i.) má EEG pravidelnejší priebeh s vyššími amplitúdami (Berger, 1929).

(2) Ukazuje sa, že vedomé stavy vyžadujú distribuovanú aktivitu v talamo-kortikálnom systéme (Baars a spol., 2003). Niektoré časti, aj

²² Alternatívne delenie ponúka napr. Block (1995), keď hovorí o vedomí typu A (access) – prístupovom, fenomenálnom, monitorovacom a type S (self). Medzi oboma taxonómiami však možno nájsť prepojenia.

²³ prípadne aj v niektorých periférnych častiach nervového systému (Pert, Schmitt, 1985), čo však teraz budem pre jednoduchosť ignorovať

keď sa priamo podieľajú na mentálnych úlohách, nemusia tvoriť súčasť tohto systému (napr. hipokampus, alebo mozoček). Iné časti môžu byť potrebné ako podporný systém na udržiavanie stavu vedomia bez toho, že by prispievali k „obsahu“ (typu) vedomia (mozgový kmeň).

(3) Senzorický vstup pri vedomí človeka vyvoláva široko distribuovanú aktivitu v mozgu, zatiaľ čo rovnaký vstup v stave nevedomia len lokálne aktivuje oblasti senzorickej kôry (Dehaene a spol., 2001). Taktiež je známe, že nové mentálne úlohy si vyžadujú viac mozgovej aktivity (rozsiahlejšie neurálne reprezentácie), ktorej objem počas automatizovania klesá a vykonávanie úlohy sa stáva nevedomým (Baars, 2002).

S týmito základnými vlastnosťami súvisia aj ďalšie podmienky pre vedomé stavy (Seth a spol., 2005), ako napr.: široký rozsah ich „obsahov“ (percepčne orientované stavy, pocity, myšlienkové procesy, atď.), limitovaná kapacita (človek si v danom čase môže byť vedomý len jednej „veci“), trvanie (rádovo zopár stoviek milisekúnd trvá, aby vznikol vedomý vnem), stabilita obsahu (napriek potenciálnej variabilnosti podnetu), uľahčenie učenia, nutnosť vedomia pri niektorých úlohách (napr. rozhodovanie sa) a ďalšie. Zaujímavou vlastnosťou primárneho vedomia je senzorické viazanie²⁴, ktoré sa vysvetľuje dvoma triedami hypotéz. Na jednej strane stoja hypotézy o synchronných osciláciách vznikajúcich v dynamicky sa tvoriacich neurálnych skupinách (napr. Crick & Koch, 1998; Singer, 1999). Na druhej strane stoja hypotézy o senzorickom viazaní prostredníctvom pozornostných procesov, exekutívnych mechanizmov a neurálnych máp (napr. Shafritz a spol., 2002). Pre obe možno nájsť empirickú podporu.

Je to možno tak, že stavy vedomia nám dávajú priestor na spoznávanie vonkajšieho sveta (prostredie) i vnútorného sveta (signály z môjho tela). Popri vedomých stavoch „o niečom“, často označovaných ako intencionálne stavy, totižto existujú aj nevedomé intencionálne stavy (napr. implicitná kognícia), ktoré taktiež zabezpečujú subjektu, aby jeho správanie bolo funkčne adaptované vzhľadom ku okolitému svetu. Je však možné, že vedomá intencionalita sa vyžaduje pri úlohách, kde by automatická (nevedomá) odozva nestačila (Dehaene & Naccache, 2001; Griffin & Speck, 2004).

²⁴ Ako zľúčiť rôzne vlastnosti tej istej entity (napr. vnímaného objektu), aby vznikol koherentný vnem.

5.2 Pohľad teórie komplexných systémov

Hľadanie kritérií pre vedomé stavy ponúka aj teória *komplexných systémov*. Účelom mozgu ako komplexného systému je zrejme jeho (evolučne vyvinutá) maximálna efektívnosť pri reprezentovaní a spracovaní informácie, čo možno považovať za kritérium konzistentné s požiadavkou na vyváženosť medzi špecializáciou a integráciou jednotlivých centier pri spracovaní informácie v mozgu (Tononi a spol., 1994).²⁵ Komplexnosť dynamiky sa dá kvantifikovať pomocou rôznych mier (napr. neurálna komplexita, integrácia informácie, či kauzálna hustota), z ktorých každá má nejaké praktické obmedzenia, avšak ich kombinácia by mohla podchytiť viacerozmernosť neurálnej zložitosti (Seth a spol., 2006). Konzistentné poznatky ponúkajú nové práce o mozgovej kôre ako o komplexnom systéme s vlastnosťami „malého sveta“, ktorý umožňuje veľmi efektívny prenos informácie v rámci systému (Sporns & Zwi, 2004). Druhou nedávno objavenou charakteristikou mozgu ako komplexného systému je jeho škálovo-nezávislá organizácia (Eguíluz a spol., 2005), ktorá napovedá, že rovnaké informačné procesy môžeme pozorovať v mozgu na viacerých úrovniach organizácie. Možnosťami metodického použitia teórie samoorganizácie a systémovej komplexnosti sa venuje Burian (2006).

Všetky uvedené vlastnosti vedomých mentálnych stavov (na globálnej úrovni) by mali predstavovať obmedzenia pri hľadaní konceptuálnych východísk pre model stelesnenej mysle.

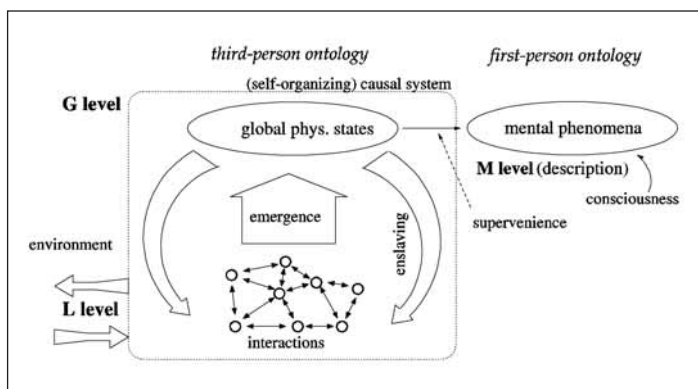
6 Úrovně opisu v modeli stelesnenej mysle

Vo Farkaš (2007) som načrtnol konceptuálny rámec pre model stelesnenej mysle z pohľadu ponúknutého vysvetlenia recipročnej kauzality (podmienka E). Schéma najjednoduchšieho takéhoto modelu je na obr. 1. Model pozostáva z *troch* (namiesto filozoficky tradičných dvoch) úrovní opisu reality (dvoch fyzikálnych a jednej mentálnej), ktoré sú nevyhnutné na pochopenie podstaty recipročnej kauzality.²⁶ Kvôli zjednodušeniu považujeme za elementárnu fyzikálnu úroveň opisu biologické neuróny a za tú najvyššiu mozog ako celok. Medzi týmito okrajovými úrovňami existuje viacero fyzikálnych úrovní organizácie, ktoré sú relevantné z pohľadu neuroanatómie a neurofyziológie (neurálne mikroobvody, neurálne zhluky, či Brodmannove kortikálne arey), aj s ich vzájomnými interakciami

²⁵ To je konzistentné s mozgom ako systémom s metastabilnou dynamikou (Bressler, Kelso, 2001, s. 26).

²⁶ Striktne povedané, z pohľadu škály má model len dve úrovne opisu (L a G), avšak pre úroveň G existujú dve vzájomne neredukovateľné *perspektívy* opisu, z ktorých obe sú nevyhnutné pre pochopenie podstaty stelesnenej mysle. Pre zjednodušenie slovníka budeme aj M chápať ako úroveň opisu.

(Churchland & Sejnowski, 1992), no v našom modeli budem od tohto faktu abstrahovať.



Obrázok 1. Náčrt formálneho modelu vzťahu myseľ-telo pozostávajúci z troch úrovni opisu. Vzťahy susedných úrovni sú kvalitatívne odlišné: len fyzikálne úrovne – lokálna L a globálna G – vzájomne kauzálne interagujú, úroveň M nemá kauzálne účinky na G – je jej (neredukovateľným) opisom IPP. Fyzikálny systém je otvorený, prijíma externé stimuly na úrovni L a späťne pôsobí na prostredie.

6.1 Fyzikálne úrovne opisu

Predpokladáme teda existenciu len dvoch interagujúcich „referenčných“ fyzikálnych úrovni, lokálnej L (časti mozgu) a globálnej G (mozog ako celok). Pritom treba zdôrazniť, že obe fyzikálne úrovne sú relatívne v tom zmysle, že fyzikálna mierka, ktorej sa budú týkať, závisí od kontextu. Napríklad, úroveň neurálnych zhlukov je globálna v kontexte úrovne neurónov, ale lokálna vo vzťahu ku kortikálnym areám. Táto abstrakcia je užitočná pre pochopenie existujúcich empirických a teoretických argumentov v prospech recipročnej kauzality, ktoré sú relevantné pre rôzne priestorové mierky (Farkaš, 2007, časť 3).

Pozícia neredukcionistického fyzikalizmu by sa dala charakterizovať takto:

Fyzikalizmus – kauzálny systém je fyzikálny, a je kauzálne uzavretý.²⁷ Kauzálne vzťahy podliehajú fyzikálnym zákonitostiam na rôznych úrovniach (od molekulárnej po systémovú), pri zjednodušenom pohľade môžeme hovoriť o úrovni neurónov. Mentálne stavy sú závislé od fyzikálnej reality, t.j. ľudská myseľ nemôže existovať bez mozgu.
Neredukcionizmus – epistemologické poznanie mysle a jej úplné vy-

²⁷ v zmysle slabej definície: „Žiaden fyzikálny jav nemá príčinu mimo fyzikálnej domény“ (Kim, 1998).

svetlenie nie je principiálne redukovateľné na opis jazykom fyzikálnej vedy (t.j. opis mozgu).²⁸

Tento filozofický smer (ktorý má svoje vnútorné členenie) nadobudol na popularite po tom, čo bol zavedený koncept *supervenencie*, podľa ktorého mentálne vlastnosti sú akousi „pridanou hodnotou“ k fyzikálnym vlastnostiam, pričom nie sú na ne redukovateľné (Davidson, 1970). Táto platforma taktiež vystihuje nesymetrickosť vzťahu medzi fyzikálnymi procesmi a mentálnymi procesmi. Existuje široká škála fyzikálnych procesov (stavov), ktoré prebiehajú v mozgu-tele, avšak iba niektoré evokujú IPP – mentálne procesy (stavy). A v rámci mentálnych stavov iba niektoré z nich sú vedomé (Dennett, 1996). Možnými kritériami pre vedomé stavy sme sa zaoberali v časti 5.

Pri súčasnej úrovni poznatkov v kognitívnej vede a príbuzných disciplínach sa mi ako ideálny prístup javí spojenie silných stránok konekcionistickej a dynamickej platformy, s cieľom splniť čo najviac podmienok pre ideálny model mysle (časť 3.2). Obe paradigmy inherentne dávajú priestor pre splnenie podmienok A a B (robustnosti a adaptívnosti), na rozdiel od symbolových systémov. Reprezentácia znalostí v kolektívnych systémoch (C) je menej transparentná, no existuje. Modelovanie procesov nižšej kognície (D1) je principiálne ľahšie prístupom „zdola“ ako „zhora“, aj keď pretrvávajúcim problémom ostáva efektívne učenie (v modeloch rekurentných neuronových sietí), hlavne v kontexte väčších modelov. Vysvetlenie procesov vyššej kognície (D2) vyzerá byť v súčasnosti schodnejšie pomocou symbolových systémov, no môže súčasne slúžiť ako prípadné vodítko pre subsymbolový prístup, ktorý by v tomto kontexte mohol byť implementáciou symbolového prístupu. Podobne by to mohlo vyzeráť v kontexte s kategóriou D3. Samozrejme, ťažko si predstaviť jeden homogénny kolektívny systém, ktorý by bol schopný vysvetliť jazykovú schopnosť u človeka. Skôr by to bol súbor prepojených systémov. Napokon, ako takýto model vysvetľuje subjektívnosť mysle (F), možno dať do súvisu s možnými neurálnymi kritériami pre vedomé stavy.

Nevyhnutnou súčasťou modelu je participácia *vonkajšieho prostredia*. Mozog-mysel si vyžaduje permanentnú interakciu s ním, a možno ešte dôležitejšie sú podnety, než samotné reakcie.²⁹ Vonkajšie prostredie je nevyhnutným zdrojom empirickej skúsenosti a z dynamického hľadiska je „partnerom“ v procese zjednávania. No kvôli fyzickej separácii od prostredia si myslím, že pri modelovaní ukotve-

²⁸ V rámci fyzikálnej domény je však model redukcionistický. Redukcionizmus sa vo vede etabloval ako najschodnejší prístup, ktorý si vyžaduje nájdenie vhodnej úrovne redukcionizmu pre daný problém (vyhnutie sa redukcii ad absurdum).

nej mysle nemusíme zachádzať za hranice ľudského tela (ako navrhuje napr. Rockwell, 2007). Myseľ sa formuje v interakcii s prostredím, jedinečne u každého jedinca, no ukotvená ostáva v tele napriek tomu, že mentálne prejavy si vyžadujú interakciu s prostredím.

Spoločenské prostredie vnáša ďalšiu dimenziu do modelu – myseľ človeka sa stáva elementom v kolektívnom systéme interagujúcich jedincov. Správanie (mentálne pochody) človeka sú výsledkom súčasného vplyvu dvoch dynamických systémov: sú emergentnými javmi v mozgu a súčasne sa stávajú podriadenými stavmi k kontexte medziľudskej interakcie. Recipročná kauzalita sa prejavuje aj na úrovni komunity.

6.2 Mentálna úroveň opisu

V náčrte modelu figuruje i mentálna úroveň opisu M, ktorá je relevantná k fyzikálnej úrovni G. Úroveň M priamo v modeli nevystupuje ako súčasť kauzálneho systému, o ktorom predpokladáme, že je kauzálne uzavretý, ale len „prostredníctvom“ úrovne G. Preto keď hovoríme o „mentálnej“ kauzalite, máme na pamäti, že „strojom“ kauzality zhora-nadol je samotný mozog, pričom mentálne stavy v mysli sú opisom globálnych stavov mozgu z 1PP. Nie je však potom opis vlastností mozgu na úrovni M redundantný (1PP), keď sa tieto nepodieľajú na riadení jeho činností? Spomeňme argumenty v prospech významu M:

(1) M predstavuje nevyhnutnú (neredukovateľnú) úroveň opisu pre kompletnejšie pochopenie mentálnych stavov. Ja ako subjekt viem „čítať“ myseľ iného subjektu, lebo dokážem správne „interpretovať“ prejavy jeho tela, nakoľko sme z „rovnakého cesta“.³⁰ Ako príklad môže slúžiť toto: Človeku, ktorý je od narodenia slepý, nemožno vysvetliť, čo je červená farba. Opis z pohľadu 3PP je preňho nepostačujúci, neposkytuje mu zážitok (aké to je). V budúcnosti by sa však tento zážitok mohol dať vyvolať správnou stimuláciou jeho mozgovej kôry (ak je funkčná). Inými slovami, 1PP je prenositeľná len prostredníctvom 3PP.

(2) M je ako jediná prístupná bežnému jazyku a jediná umožňuje efektívnu verbálnu komunikáciu medzi subjektmi. Subjekt vlastne ani nemá na výber, pretože úroveň G (aktivita jeho vlastného tela-mozgu) mu je neprístupná. Súčasne tým umožňuje subjektu figurovať v kontexte populácie, ktorú môžeme chápať ako vyššiu úroveň organizácie

²⁹ Dlhodobejší stav senzorickej deprivácie je pre človeka neprirodený a vedie k destabilizácii mysle, zatiaľ čo dlhodobejšia nečinnosť niektorým jedincom neškodí ☺.

³⁰ a ktoré by boli rovnaké (alebo aspoň podobné) u mňa, keby som pozorovanú činnosť vykonával – ja)

s možnými komplexnými prejavmi kolektívneho správania. Aj keby sme v budúcnosti našli presné neurálne koreláty všetkých mentálnych stavov, ich opis by bol kvôli ich zložitosti prakticky nepoužiteľný (pozri aj súvis jazyka a vedomia v Beňušková, 2005 a v Dennett, 1996).

(3) Vedomé stavy súvisiace s niektorými stavmi v mozgu (časť stavov v M) sú možno emergentným fenoménom. Argument o zombiách (hypotetických tvoroch – ľudských kópiách s absenciou vedomia), podporujúci nevysvetliteľnosť „ťažkých“ problémov vedomia pomocou vedeckých metód je logicky možný, no možno nerealistický.³¹ Sám Chalmers (1996) pripúšťa, že existencia zombií je málo pravdepodobná (s. 96). Z pohľadu nášho modelu vedomé stavy majú mentálnu kauzálnu silu prostredníctvom úrovne G, pričom si uvedomujeme príčinné vzťahy. Máme pocit, že *my* sa rozhodujeme, máme pocit slobodnej vôle. Súčasne však zrejme i nevedomé mentálne stavy majú kauzálnu silu („nadol,“ resp. i v rámci kauzálnej domény), pričom sprievodné účinky sú mimo vedomia (človek si nie je vedomý príčin svojho rozhodovania). Absencia participácie vedomia je typická pre rýchle konanie, napr. v športe (hokejista si až dodatočne uvedomí a precitne, aký dal krásny gól), ale možno aj pri pomalších procesoch. To preto, že sme pod vplyvom mnohých podnetov (mimo tela i v rámci neho), mozog všetky spracováva, no iba niektoré procesy sa dostanú na vedomú úroveň.

(4) Vedomie (a teda špecifické stavy/procesy mozgu) sú možno žiaducim, evolučne vyvinutým fenoménom pre človeka ako sociálne a emocionálne založený živočíšny druh (Kováč, 2003). O emóciách som sa v príspevku vyjadroval len okrajovo, ale nepochybne limbický systém zohráva významnú úlohu v kognitívnych procesoch (LeDoux, 1996). Mohli sme ho chápať ako modulačný systém, ktorý na úrovni tela interaguje s neurálnym systémom.

(5) Vedomé stavy sú dôležité, pretože výrazne podporujú učenie, t.j. aktualizáciu dlhodobej pamäti (Seth a spol., 2005). Toto tvrdenie nie je v konflikte s poznatkami o implicitnom učení (Cleeremans, 1997), ktoré nastáva pri vedomí, ale bez prítomnosti úmyslu (intencie) učiť sa.³² To napovedá o mechanizmoch učenia aj bez prítomnosti vedomia.

³¹ Ak je niečo „logicky“ možné, predstaviteľné, nemusí mať ešte oporu vo fyzikálnej realite. Napr. zdravý rozum nám umožňuje predstaviť si teleso letiace väčšou rýchlosťou ako rýchlosť svetla (generalizácia), aj keď v skutočnosti to nie je možné.

³² Napríklad, subjekt je inštruovaný, aby sa sústredil na nejakú úlohu, no potom sa testuje na inej úlohe, pričom sa zistí, že sa ju tiež naučil (vo väčšej alebo menšej miere).

7 Zhrnutie

V príspevku som uviedol viacero argumentov, prečo si myslím, že pohľad z perspektívy kolektívnych systémov so samoorganizáciou má najbližšie k pochopeniu ontologickej podstaty mysle, ako nehmotnej entity ukotvenej vo fyzikálnej realite (mozgu-tele). Takýto formálny pohľad navyše umožňuje vysvetľovať mechanizmy, o ktoré sa opierame pri mentálnych procesoch, pričom tieto mechanizmy spolu s architektúrou kolektívneho systému implikujú vlastnosti pre ideálny model mysle. Fyzikálne ukotvenie mysle v mozgu-tele ide ruke v ruke s hľadaním neurálnych korelátov, ktoré sú predmetom exaktného experimentálneho výskumu. Subjektívnosť ľudskej mysle možno vnímať ako evolučné riešenie, zabezpečujúce efektívnu existenciu človeka v dynamickom prostredí. Toto riešenie možno bude a možno nebude v budúcnosti replikovateľné na inej fyzikálnej báze, pričom súčasná neuroveda napovedá, že skôr nie. Či sa jazýčky váh preklopia, a či to budeme vedieť rozhodnúť, ukáže (snáď) iba budúcnosť.

PodĎakovanie: Tento príspevok vznikol za podpory grantovej agentúry VEGA v rámci grantových úloh 1/3612/06 a 2/7087/27.

Literatúra

Aizawa K.: *The Systematicity Arguments*. Kluwer Academic, Dordrecht 2003.

Atkinson A., Thomas S., Cleeremans A.: Consciousness: mapping the theoretical landscape. *Trends in Cognitive Sciences*, 4 (2000) 372-382.

Baars B.: The conscious access hypothesis: Origins and recent evidence. *Trends in Cognitive Sciences*, 6 (2002) 47-52.

Baars B., Banks P., Newman J. (eds.): *Essential Sources in the Scientific Study of Consciousness*. Cambridge, MA: MIT Press 2003.

Barsalou L.: Perceptual symbol systems. *Behavioral and Brain Sciences*, 22 (1999) 577-609.

Bell A.: Levels and loops: The future of artificial intelligence and neuroscience. *Philosophical Transactions of the Royal Society London B*, 354, 2013-2020, 1999.

Beňušková Ľ.: Kde sa jazyk stretáva s vedomím. In: *Jazyk a kognícia*, Kalligram, Bratislava, 235-261, 2005.

Berger H.: Über das Elektroenkephalogramm des Menschen. *Archiv für Psychiatrie und Nervenkrankheiten*, Berlin, 87 (1929) 527-570.

Block N.: On a confusion about a function of consciousness. *Behavioral and Brain Sciences*, 18 (1995) 227-287.

Burian J.: Samoorganizace a kognice. In: *Mysel, inteligencia a život*. Vydavateľstvo STU v Bratislave, 2007, 173-190.

- Calvin W.H., Ojemann G.A.: *Conversations with Neil's Brain. The Neural Nature of Thought and Language*, Addison-Wesley 1994.
- Clark A.: *Mindware: An Introduction to the Philosophy of Cognitive Science*. Oxford University Press 2001.
- Chalmers D.: *The Conscious Mind*. NY: Oxford University Press 1996.
- Chalmers D.: First-person methods in the science of consciousness. *Arizona Consciousness Bulletin*, 1999, <http://consc.net/papers/firstperson.html>.
- Churchland P.: *The Neurocomputational Perspective*. Cambridge, MA: Bradford Books 1989.
- Cleeremans A.: Principles of implicit learning. In D. Berry (Ed.) *How implicit is implicit learning?* Oxford: Oxford University Press (1997) 196-234.
- Cleeremans A.: Computational correlates of consciousness. *Progress in Brain Research* 150 (2005) 81-98.
- Crick F., Koch C.: Consciousness and neuroscience. *Cerebral Cortex*, 8 (1998) 97-107.
- Davidson D.: *Mental Events*. Reprinted in *Essays on Actions and Events*. Oxford: Clarendon Press 1970.
- Dehaene S., Naccache L., Cohen L., Bihan D., Mangin J., Poline J., Riviere D.: Cerebral mechanisms of word masking and unconscious repetition priming. *Nature Neuroscience*, 4 (2001) 752-758.
- Dehaene S., Naccache L.: Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. *Cognition* 79, (2001) 1-37.
- Dennett D.: *The Intentional Stance*. Cambridge, MA: MIT Press 1987.
- Dennett D.: *Kinds of Minds: Towards an Understanding of Consciousness* 1996. Preklad: Peregrin J.: Academia Praha 2004.
- Doya K.: What are the computations of the cerebellum, the basal ganglia, and the cerebral cortex? *Neural Networks*, 12 (1999) 961-974.
- Eguíluz V. a spol.: Scale-free brain functional networks. *Physical Review Letters* 94 (2005) 018102.
- Farkaš I.: Konekcionistické modelovanie jazyka. In: *Jazyk a kognícia*, Kalligram, Bratislava, 262-305, 2005.
- Farkaš I.: Samoorganizácia ako hybná sila dynamických vzorcov aktivít v mozgu a mysli. In: *Kognície a umelý život VI*, Opava: Slezská univerzita, 143-148, 2006.
- Farkaš I.: Hľadanie kauzálnych vzťahov v probléme mysle a tela z pohľadu neredukcionistického fyzikalizmu. In: *Mysel, inteligencia a život*. Vydavateľstvo STU v Bratislave, 2007, 3-16.
- Fodor J.: *The Modularity of Mind*. Cambridge, MA: MIT Press 1983.
- Fodor J.: *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. Cambridge, MA: MIT Press 1987.
- Fodor J., Pylyshyn Z.: Connectionism and cognitive architecture: A critical analysis. *Cognition*, 28 (1988) 3-71.

- Friedenburg J., Silvermann G.: *Cognitive Science: An Introduction to the Study of Mind*. London, UK: Sage Publications 2006.
- Gärdenfors P. Cued and detached representations in animal cognition. *Behavioural Processes*, 36 (1996) 263-273.
- Griffin D., Speck G.: New evidence of animal consciousness. *Animal and Cognition*, 7 (2004) 5-18.
- Havel I.: Přirozené a umělé myšlení jako filozofický problém. V kniže Mařík V. a kol.: *Umělá inteligence (3. díl)*. Academia Praha, 2001, 17-75.
- Havel I.: Prožívání epizodických situací. In: Kvasnička V. a spol. (ed.): *Mysel, inteligencia a život*. Vydavateľstvo STU v Bratislave, 2007, 27-70.
- Harnad S.: The symbol grounding problem. *Physica D*, 42 (1990) 335-346.
- Hornik K., Stinchcombe M., White H.: Multilayer feedforward networks are universal approximations. *Neural Networks*, 2 (1989) 359-366.
- Hurley S., Chater N. (Eds.): *Perspectives on Imitation: From Neuroscience to Social Science* (Vol. 1, 2), Cambridge, MA: MIT Press, 2005.
- Izhikevich E.: Which model to use for cortical spiking neurons? *IEEE Transactions on Neural Networks*, 15 (2004) 1063-1070.
- Jedlička P., Beňušková L., Macáková J., Ostatníková D.: Molekulové mechanizmy učenia a pamäti. In: Hulín I. (ed.) *Patofyziológia* (6. vyd.) Slovak Academic Press, Bratislava, (2002) 1183-1199.
- Johnson-Laird P.: *Mental Models*. Cambridge University Press 1983.
- Hutchins E.: *Cognition in the Wild*. Cambridge, MA: MIT Press 1995.
- Kaczmarek L., Levitan I. (eds): *Neuromodulation: The Biochemical Control of Neuronal Excitability*. NY: Oxford University Press 1987.
- Kelso, J.A.S.: *Dynamic Patterns*. Cambridge, MA: MIT Press 1995.
- Kim J.: *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*. Cambridge MA: MIT Press 1998.
- Kosslyn S.: *Image and Brain: The Resolution of the Imagery Debate*. Cambridge, MA: MIT Press 1994.
- Kováč, L.: Ľudské vedomie je produktom evolučnej eskalácie emocionálneho výberu. In: *Kognícia a umelý život III*. Elfa: Košice (2003) 21-39.
- Kráľ A., Hulín I.: Neurofyziologické predpoklady vedomia. *Psychiatria*, 8 (2001) 86-98.
- Kvasnička V.: Kognitívna veda, umelá inteligencia a teória ohraničenej racionality. In: *Kognice a umelý život VI*, Opava: Slezská univerzita, 233-252, 2006.
- LeDoux K.: *The Emotional Brain*. NY: Touchstone 1996.
- Marr D.: *Vision*. San Francisco, CA: W.H. Freeman 1982.
- Mayberry M.: Incremental nonmonotonic parsing through semantic self-organization. *Doctoral thesis*. Department of Computer Sciences, University of Texas at Austin 2003.

- Metzinger T. (ed.): *Neural Correlates of Consciousness: Empirical and Conceptual Questions*. Cambridge, MA: The MIT Press 2000.
- Newell A., Simon H.A.: Computer science as empirical enquiry. *Communications of the ACM*, 19 (1976) 113-126. In: *Mind Design II*, J. Haugeland (ed.), 81-110. Cambridge, MA: MIT Press 1997.
- O'Reilly R., Munakata Y.: *Computational Explorations in Cognitive Neuroscience: Understanding the Mind by Simulating the Brain*. MIT Press 2000.
- Pert C., Schmitt F.: Neuropeptides and their receptors: A psychosomatic network. *Journal of Immunology*, 135 (1985) 820-826.
- Port R., van Gelder T. (Eds.): *Mind as Motion: Explorations in the Dynamics of Cognition*. Cambridge, MA: MIT Press 1995.
- Pstružina K.: Abstrakce, komparace a generalizace jako základní operace lidského myšlení. In: *Myseľ, inteligencia a život*. Vydavateľstvo STU v Bratislave, 2007, 111-117.
- Pylyshyn Z.: *Computation and Cognition*. MIT Press 1984.
- Ramsey W., Stich S., Garon, J.: Connectionism, eliminativism and the future of folk psychology. In: *Philosophy and Connectionist Theory*. Hillsdale, NJ: Lawrence Erlbaum (1991) 199-228.
- Rizzolatti G. Fadiga L., Gallese V., and Fogassi L.: Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, 3(2), (1996) 131-141.
- Rockwell T.: *Neither Brain nor Ghost: A Nondualist Alternative to the Mind-Brain Identity Theory*. Cambridge, Mass.: MIT Press, 2007.
- Rohde, D. A connectionist model of sentence comprehension and production. *Doctoral thesis*, Carnegie Mellon University, Pittsburgh, PA 2002.
- Rohde D., Plaut D.: Connectionist models of language processing. *Cognitive Studies*, 10 (2003) 10-28.
- Searle J.: Consciousness. *Annual Reviews of Neuroscience* 23 (1999) 557-578.
- Seth A., Baars B., Edelman D.: Criteria for consciousness in humans and other mammals. *Consciousness and Cognition*, 14 (2005) 119-139.
- Seth A., Izhikevich E., Reeke G., Edelman G.: Theories and measures of consciousness: An extended framework. *Proceedings of the National Academy of Sciences*, 103 (2006) 10799-10804.
- Shafritz K., Gore J., Marois R.: The role of the parietal cortex in feature binding. *Proceedings of the National Academy of Sciences*, 99 (2002) 10917-10922.
- Siegelmann H., Sontag E.: Turing computability with neural nets. *Applied Mathematics Letters*, 4 (1991) 77-80.
- Singer W.: Neuronal synchrony: A versatile code for the definition of relations? *Neuron*, 24, (1999) 49-65.
- Spitzer M.: *The Mind within the Net: Models of Learning, Thinking and Acting*. Cambridge, MA: MIT Press 1999.
- Sporns O., Zwi J.: The small world of the cerebral cortex. *Neuroinformatics*, 2 (2004) 145-162.

- Sternberg R.: *Cognitive Psychology* (3. vyd.). Thomson Wadsworth 2003.
- Šefránek J.: Kognitívna teória usudzovania. In: *Mysel, inteligencia a život*. Vydavateľstvo STU v Bratislave, 2007, 233-239.
- Thelen E., Smith L.: *A Dynamic Systems Approach to the Development of Cognition and Action*. Cambridge, MA: MIT Press 1994.
- Tononi G., Sporns O., Edelman G.: A measure for brain complexity: relating functional segregation and integration in the nervous system. *Proceedings of the National Academy of Science*, 91 (1994) 5033-5037.
- van Gelder T.: Compositionality: A connectionist variation on a classical theme. *Cognitive Science*, 14 (1990) 355-384.
- van Gelder T., Port R.: It's about time: An overview of dynamical approaches to cognition. In: *Mind as Motion: Explorations in the dynamics of cognition*. Cambridge, MA: MIT Press, 1995, 1-44.
- Varela F.J., Thompson E., Rosch, E. *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge, MA: MIT Press 1991.
- Vavřečka M.: Ukotvení symbolů v kontextu nulového sémantického závazku. In: *Kognícia a umelý život VI*, Opava: Slezská univerzita, 2006, 401-411.
- Wiedermann J.: Mirror neurons, embodied cognitive agents and imitation learning. *Computing and Informatics*, 22 (2003) 545-559.
- Ziemke T.: What's that thing called embodiment? *Proceedings of the 25th Annual Meeting of the Cognitive Science Society*. Lawrence Erlbaum 2003.
- Zwaan R., Radvansky A.: Situation models in language comprehension and memory. *Psychological Bulletin*, 123 (1998) 162-185.

Turingův stroj, test, tělo a jablko

Jana Horáková¹ a Jozef Kelemen²

1 Prolog

Historii lidstva se kutálí jablko. Jablko poznání a hříchu. Kdysi na úsvitu dějin byl první test. Eva nabídla jablko Adamovi, on se zakousl... důsledky neseme dodnes. Pak bylo jablko, co padlo Newtonovi na hlavu, a nyní vládeme poznáním některých tajemství fyziky. V dětství jsme v představách prožívali osudy princezny, která ochutnala jablko připravené pro hřích. Pro vraždu. Neumřela, pouze dlouho, dlouho spala. Prožívali jsme i osudy prince, který ji našel a svou láskou probral k novému životu. Všechno jsou to pohádky, řeknete si.

A co například Turingův test, jeho poznání, jeho stroj, jeho láska a jeho tělo? No a co jeho jablko...? Jeho lidský osud to všechno propojuje. Jak uvádí David Leawitt (Leawitt, 2007, s. 237), coby student si mladý Turing ještě několik týdnů po shlédnutí filmu *Sněhurka a sedm trpaslíků*³ prozpěvoval na chodbách cambridgské King's College:

*Jablko nořím do jedu
Na svět spavou smrt vedu...⁴*

Jak to asi s ním šlo dál...?

2 Turing

Alan Mathison Turing se narodil 23. června 1912 v londýnské části Peddington a zemřel 7. června 1954 ve Wilmslow v Cheshiru. Byl *geniálním matematikem* a také *homosexuálem* v Británii druhé poloviny minulého století.⁵ Hned tři životní komplikace mu osud dopřál! Zda, v čem a nakolik tyto tři vlastnosti – být geniálním, být matematikem

¹ Ústav hudební vědy, Filozofická fakulta Masarykovy university, A. Nováka 1, 60200 Brno, ČR, e-mail: horakova@phil.muni.cz.

² Ústav informatiky, Filozoficko-přírodovědecká fakulta Slezské university, Bezručovo nám. 13, 74601 Opava, ČR, e-mail: kelemen@fpf.slu.cz a Vysoká škola manažmentu, Panónska cesta 17, 851 04 Bratislava, SR.

³ Tento film (v originálu *Snow White and the Seven Dwarfs*) byl prvním celovečerním animovaným filmem v produkci Walta Disneyho, prvním celovečerním animovaným filmem, natočeným v anglickém jazyce vůbec a prvním animovaným filmem oceněným Americkou filmovou akademií známou soškou Oscara. Do distribuce byl uveden v roce 1937, premiéra v USA byla 21. prosince 1937. V Británii, kde ho shlédl i Alan Turing, se film začal promítat 12. března 1938.

⁴ V originálu ve filmu *Snow White*: „Dip the apple in the brew / Let the Sleeping Death seep through“. České přebásnění jsme převzali z (Leawitt, 2007).

⁵ Sadie Plant v knize *Zeros and Ones: Digital Women and New Technoculture* (1997) spekuluje o tom, že duhové jablko v logu firmy Apple je poctou Turingovi. Jde však asi pouze o legendu, vždyť duhová vlajka homosexuálů byla navržena Gilbertem Bakkerem až v roce 1978, kdy Apple už své logo mělo.

a být homosexuálem – ovlivňovaly jeho osobnost a osud, je v současnosti předmětem různých studií, debat i spekulací. Některé z nich v této stati připomeneme (a možná přidáme i nějaké vlastní).

Byl nadaným dítětem, tak trochu podivínským. Vyberáme ukázkou ze začátku 4. scény divadelní hry Valerie Patery *Alanovo jablko* (Patera, 2004):

Alan a Christopher ve školní knihovně.

Christopher: Oh, Turing, co tu v knihovně děláš? Ještě nikdy jsem tě tu nepotkal...

Alan: Jo, to máš pravdu, jenom jsem si pomyslel, že je to tu dobré místo k přemýšlení...

Christopher: To je...

Alan: Jo, to je...

(Po chvílce plachého zaváhání)

Vypočítal jsem π na šestatřicet desetinných míst...

Turingovi to s přemýšlením a počítáním šlo i po absolutoriu. Vědecké dílo, které po sobě zanechal, ovlivnilo dokonce směřování vědy a techniky 20. století velice výrazně. Dnes snad již není pochyb ani o tom, že digitální informační technologie – jedna z důsledků Turingovy vědecké činnosti – mnohostranně a hluboce ovlivňují i naši obecnou kulturu a utváření i fungování společnosti, ve které žijeme, tedy nás samé.⁶ Současně však platí i to, že lidé si vždy přivlastňují technologie nepředvídatelným způsobem. Do jaké míry se však odrážejí naše touhy a obavy, sny a viny v předpokladech a očekáváních, které s novými technickými aparáty a jejich využíváním spojujeme?

Tuto otázku většinou řešíme v souvislosti se vzájemným působením vědeckých objevů a technických vynálezů na jedné straně a oblastí umění, filozofie nebo vědecké fantastiky na straně druhé. My však nyní budeme spekulovat o tom, jakou roli mohl sehrát osud vědce Alana Turinga v případě vzniku a pozdějších dopadů jeho hypotézy „inteligentního stroje“ a ve výstavbě jeho argumentace založené na určitém oddělení mysli a těla, která tuto hypotézu podpořila.

Řeč bude tedy nejenom o Alanu Turingovi jako o jednom z nejvýznamnějších vědců 20. století. Bude i o jeho lidském osudu a o seipjatosti tohoto osudu s dílem, které dalo 20. století jednu z jeho určujících technologických, vědeckých, uměleckých a obecně společenských orientací.

⁶ Podrobněji jsme se o těchto změnách zmínili v (Kelemen, 2007).

3 Turingův stroj

Turing se za 2. světové války vyznamenal tím, že sehrál klíčovou roli při rozluštění způsobu kódování německého šifrovacího stroje *Enigma* a výrazně tak pomohl k vítězství spojenců⁷: V letech 1939-1940 spolu s matematikem Gordonem Welchmanem, taky z Cambridge, zkonstruovali stroj, který záhy získal pojmenování *britská bomba*. Jeho základní vlastností bylo, že dokázal prolomit každou správu, šifrovanou strojem *Enigma*. Jeho konstrukci pomohlo, že "hardware" tohoto šifrovacího stroje byl Turingovi a jeho kolegům z Bletchley Parku, kde sídlila příslušná složka britské armády, znám a textový, asi dvacetipísmenný „tahák“, potřebný k prolomení *Enigmou* aktuálně používaného šifrovacího klíče, mohl být získán vcelku spolehlivě analýzou vhodně vybraných šifrovaných zpráv nepřítele⁸.

Ještě před válkou však Turing získal odborný věhlas svým návrhem teoretického, matematického konceptu stroje, který dokázal zpracovávat informace (vykonávat algoritmizované výpočty), tzv. *Turingova stroje* (Turing, 1936), a stal se tak zakladatelem počítačové vědy. Patera si to u Turingů po návrhu stroje představila následovně (úryvek je z 11. scény její již zmíněné hry):

U Turingů: Julius (Alanův otec), Ethel (Alanova matka) a John (Alanův bratr) sedí kolem jídelního stolu. Alan přichází lapajíc po dechu, v nedbale zapnutém saku, ve dvou různých botách a neučesaný.

....

Alan: Začal jsem se zabývat jedním teoretickým strojem. A oni ho nazvali Turingův stroj!

Julius: (*vypadá hrdě*) Oh, to zní skvěle, to zní skutečně skvěle: Turingův stroj.

Athel: Máme to v rodině, je to v našich genech. Vždyť náš blízký příbuzný, George Stoney, ten navrhl elektron!

Alan: Maminko, stokrát jsem ti už říkal, že on elektron nenavrhl, protože elektron již existoval. On ho objevil a pojmenoval...

⁷ Jak ukazuje film *Enigma* režiséra Michaela Apteda z roku 2001 natočený podle scénáře Toma Stopparda, který ho napsal podle stejnojmenného románu Roberta Harrise. Turingem je ve filmu inspirovaná postava Toma Jericha, kterou hraje Dougray Scott. Koproducentem filmu byl i Mick Jagger ze skupiny Rolling Stones. Někteří kritici vytýkají tvůrcům eliminaci Turingovy homosexuality, dokonce, vzhledem k tomu, že Tom Jericho prožívá ve filmu hned dva heterosexuální vztahy, překrucování historické skutečnosti, což je ovšem v současné kinematografii běžným (protože komerčně úspěšným) postupem.

⁸ Zájemce o tuto dobrodružnou část Turingova života a působení o řadu zajímavých dokumentů a jiných informací odkazujeme na www.turing.org.uk/sources/mathenigma.html.

Ethel: To na věci nic nemění! Musíš o tom říct těm tvým Američanům!

Alan: (*Ironicky*) Co zítra ráno udělám jako první bude, že jim napíšu dopis!

(*Julius a John se na sebe usmějí*) ...

Brněnskému rodákovi Kurtovi Gödelovi sa jako prvnímu podařilo specifikovat v jistém přesném matematickém smyslu třídu vypočitatelných funkcí. Nazval prvky této skupiny funkcí rekurzivními funkcemi (Gödel, 1931). Alonso Church ukázal (Church, 1936), že třída, kterou definoval on jiným způsobem, než jakým postupoval Gödel, za pomoci tzv. λ -kalkulu, vymezuje stejnou množinu funkcí – tzv. λ -definovatelných funkcí, jakou se podařilo vymezit Gödelovi. Když v návaznosti na výše zmíněné Gödelovi a Churchovi články přemýšlel Turing o tom, jak může vypadat univerzální výpočetní zařízení, které „mechanicky“ vypočítává hodnoty matematicky korektně definovatelných funkcí, vydal se najít pro výpočet hodnot rekurzivních funkcí co nejjednodušší zařízení. Opřel se o analogii takovéhoho výpočetního zařízení s psacím strojem, u kterého si představil dostatečně dlouhou (potenciálně oboustranně nekonečně dlouhou) pásku, nejenom s možností psát na ni, nýbrž i mazat z ní již napsané znaky. Matematicky přesný popis *Turingova stroje*, který z těchto počátečních představ vzešel (Turing, 1936), je dnes k nalezení v každé spolehlivé učebnici teoretické informatiky.

Tématem Turingova průkopnického článku byla vypočitatelné čísla (angl. *computable numbers*). *Tyto čísla mohou být stručně popsány*, píše Turing hned v první větě svého článku, *jako reálná čísla, jejichž vyjádření číslicemi desítkové soustavy je vypočitatelné konečnými prostředky. I když jsou výlučným předmětem tohoto článku vypočitatelná čísla*, pokračuje Turing, *je téměř stejně lehké definovat a zkoumat vypočitatelné funkce celočíselných, reálných, nebo vypočitatelných proměnných, vypočitatelné predikáty, apod. (...) Ve smyslu mé definice je číslo vypočitatelné, dá-li se jeho dekadické vyjádření napsat strojem*. Pak samozřejmě následuje již formalizovaný matematický text, pro jehož čtení je třeba mít určitou matematickou přípravu.

K návrhu *Turingova stroje* (v Turingově původní terminologii z jeho článku z roku 1936 tzv. *a-stroje*; trochu uniká obecné odborné pozornosti, že v tom článku Turing naznačil i možnost jiného chápání abstraktních strojů) došlo, jak jsem to již výše naznačili, z vnitřních důvodů rozvoje matematiky. Ta se dostala začátkem minulého století do stádia, kdy se jako velice naléhavou stala otázka efektivní vyčísli-

telnosti funkcí. Všichni jsme se kdysi učili, že funkce jsou jakési přešné předpisy přiřazující hodnotám jejich proměnných určité výsledné hodnoty. Funkce označená třeba + přiřazuje každé dvojici čísel, třeba např. konkrétním číslům, 2000 a 8, jejich součet, v našem případě číslo $+(2000, 8) = 2008$ (což jsme zvyklí psát jako $2000 + 8 = 2008$). Je však pro každou myslitelnou funkci možné přesně definovat konstruktivní posloupnost základních operací *Turingova stroje*, kde vykonáním operací v dané posloupnosti se zaručeně dopracuje stroj k výsledku? Existuje ke každé matematicky definovatelné funkci algoritmus, který vyčíslí jeho výsledek pro všechny hodnoty jeho definičního oboru? To byly – velice zhruba řečeno – otázky, které trápily matematiky toho období.

Dvěma implicitními ale – jak se ukázalo koncem 20. století – podstatnými znaky *Turingova stroje* jsou (1) dokonalá *izolovanost* (*uzavřenost*) *Turingova stroje* od jeho vnějšího prostředí (se znaky na pásce se nic neděje z jiné příčiny než z té, že je příslušná změna naprogramována v jeho řídicí jednotce) a (2) *perfektnost Turingova stroje* – co *Turingův stroj* má vykonat, to vždy vykoná, a nikdy nevykoná nic, co mu jeho program nepřikazuje. Oba tyto požadavky jsou naprostou samozřejmostí, vyžadujeme-li od *Turingova stroje* to, na co byl navržen, tj. vykonávání algoritmických výpočtů inspirovaných potřebami matematiky. Jsou však velice problematické, přemýšlíme-li o nich jako o vlastnostech libovolného reálně v našem materiálním světě pracujícího (přirozeného nebo umělého) zařízení.

Co by se však stalo, kdybychom předpokládali, že právě idealizovaná představa stroje mající znaky (1) a (2) byla Turingu blízká? Blízká ne pouze jako matematikovi, nýbrž i jako člověku? Andrew Hodges v Turingově biografii (Hodges, 1983) tvrdí, že stroje, komunikující pouze pomocí dálkopisu, byly ideálem Turingova vlastního stylu komunikace, jeho vlastního bytí ve společnosti, jeho života, ve kterém mohl být ponechán osamocen ve svém pokoji a s vnějším světem by komunikoval pouze prostřednictvím racionálních, logických argumentů.

4 Turingův test

Co všechno nám *Turingův stroj* umožňuje chápat jako vypočitatelné? Třeba vše, co děláme a co chápeme jako důsledek naší lidské dovednosti, naší lidské inteligence, je ve své podstatě výsledkem výpočtů, které je schopen provádět i *Turingův stroj*? Jinak řečeno: Mohou počítače myslet jako lidi? Tato otázka nemohla neoslovit v rámci evropské kultury, kde je již staletí implicitně nebo explicitněji přítomná, vzdělané a přemýšlivé lidi, kteří byli ve druhé polovině minulého století konfrontováni s existencí *Turingova stroje* a jemu podobných ab-

straktních modelů výpočtů, anebo se schopnostmi prvních reálných hardwarově uskutečněných počítačů. Turing patřil do obou skupin. Snad to bylo důvodem, proč byl přizván na diskusi, kterou bychom mohli pokládat za začátek *kognitivní vědy* v Evropě, na besedu organizovanou na Manchesterské univerzitě filozofkou Dorothy Emmet, o možnosti počítačů myslet. Beseda sa konala 27. října 1949.

O potenciálu, který se v počítačích a ve vynořující se možnosti konstruovat sítě umělých neuronů skrývá, se přemýšlelo i za Atlantikem. Jak se v dopisu Norbetu Wienerovi z 29. listopadu 1946 zmínil John von Neumann, inspiraci v Turingově stroji nacházeli podle něho i Warren McCulloch a Walter A. Pitts, když tvořili svůj formalizovaný model neuronu, i když je pravdou, že ve svém fundamentálním článku (McCulloch, Pitts, 1943) Turinga přímo necitují.

S Claudem E. Shannonem, zakladatelem moderní kvantitativní teorie informace, který v té době pracoval v Bellových výzkumných laboratořích, diskutoval Turing o těchto možnostech osobně, když se s ním v roce 1937, v době své návštěvy Spojených států, setkal. Shannon byl nadšen jeho článkem o *a-stroji* a taky představou, že by činností stroje mohl být napodoben lidský mozek. Domníval se totiž, že pak by mohl takový stroj myslet. Turing si pak jednou zavzpomínal, jak mu Shannon říkal, že by chtěl do takového mozku dostat nejenom data, nýbrž i prvky kultury. Chtěl by, aby takový stroj hrál šachy a skládal hudbu... Patera si představila, že část jejich rozhovorů mohla vypadat i takhle:

Amerika, Alan a Shannon jedí v noční laboratoři; vedle je šachovnice. Alan se dívá na cibuli na Shannonově talíři a píchá do ní vidličkou.

Alan: Cibule, jo, byla to právě cibule, co mi dala vhlédnout do lidské mysli... cibule, jejíž vrstvy mohou být odstraňovány (*prsty opatrně oddělí jednu slupku*) ... Uvažujeme-li o funkcích mysli nebo mozku, nacházíme určité operace, které mohou být vysvětleny v ryze technických pojmech. To však neodpovídá „skutečné“ mysli; je to druh slupky, kterou musíme odstranit, abychom ji našli. Jakmile ji však odstraníme, najdeme další slupku k odstranění a pak další a další. Dostaneme se takhle ke “skutečné” mysli, anebo pouze ke slupce, kde uvnitř již nebude nic?

...

Alan: Víte Shannone, to co já hledám, je nějaký princip, teorie, obecné pravidlo, které může být odvozeno z mé hry. Jsem si jist, že pro šachy může být sestavena „definitivní“ metoda, tedy stroj. To ještě nemusí samozřejmě znamenat vytvoření skutečného stroje, ale třeba

jenom série pravidel, které může sledovat i hráč „bez mozku“, tedy jakési tabulky instrukcí, šachového programu.

Shannon: Chlapče, jsme na stejné vlnové délce! Čas od času vyhrávám partii jednoduše tak, že aplikuji základní pravidla... Hráč často uvidí fantastické tahy, které závisí od toho, jak potáhne protivník, ale...
(*Vykoná svůj tah*)

Alan: Jo, logika minimaxu; výběr nejlepšího ze zlých tahů... No jo, ale kde si myslíte, že potáhnu tímto střelcem? (*Udělá brilantní tah*) Šachmat!

Turing nejel na manchesterskou univerzitní diskusi o potenciálních schopnostech strojů intelektuálně nepřipraven. Po diskusi však měl (asi právem) dojem, že se mu nepodařilo všechny účastníky, mezi nimi již tehdy známého a uznávaného filozofa Michaela Polanyiho, přesvědčit o tom, že stroje jednou budou moci myslet. Proto se rozhodl vyjádřit svůj názor konzistentněji, v podobě článku, psaného se zvláštním zřetelem na čtenáře-filozofy. Napsal a v roce 1950 v prestižním filozofickém časopisu *Mind* publikoval článek *Computer Machinery and Intellingence* (Turing, 1950), ve kterém navrhl experiment, založený na principech *imitační hry*, později známý jako *Turingův test*:

Hru hrají tři hráči, muž (A), žena (B) a moderátor (C), na jehož pohlavi nezáleží. Moderátor je v jiné místnosti než ostatní osoby. Cílem hry moderátora je určit, která ze dvou osob je muž a která žena. Zná je pouze jako hráče X a hráče Y a na konci hry řekne X je A a Y je B nebo X je B a Y je A. Moderátor může hráči A i hráči B klást otázky. Může se například zeptat:

C: Řekne mi X, jak má dlouhé vlasy?

Předpokládejme nyní, že X je skutečně A. Musí odpovědět. Cílem hry hráče A je, aby se hráč C při identifikaci zmýlil. Jeho odpověď proto může být:

Mám vlasy ostříhané nakrátko, ne víc než přibližně 9 palců dlouhé. Odpovědi by měly být napsané, nejlépe strojem. Ideální by byla komunikace prostřednictvím dálkopisu.

Cílem hry hráče B je pomáhat moderátorovi. Jeho nejlepší strategií asi je, pokud dává pravdivé odpovědi. Ke svým odpovědím může přidat informace jako Já jsem žena, neposlouchejte ho!, tím však ničeho nedosáhne, neboť muž může udělat to samé.

Teď se zeptáme, co se stane, pokud v této hře hráče A nahradíme strojem. Bude se moderátor při svém rozhodování mýlit stejně často, jako když hraje muž a žena? (...)⁹

⁹ Výše uvedený citát je překladem provedeným na základě slovenského překladu (Turing, 1950) v (Gál, Kelemen, 1992, s. 18).

Turing v uvedeném testu zřetelně odděluje fyzické schopnosti člověka od intelektuálních, postupuje tedy velice podobně, jako když v případě *Turingova stroje* plně zanedbal eventuální výnam toho, co získalo po konstrukci reálných počítačů název *hardware*. Vše, kromě křišťálové čisté inteligence, musí být z testu odstraněno: *Naše formulace problému neumožňuje moderátorovi vidět nebo se dotýkat účastníků hry ani slyšet jejich hlasy*, zdůrazňuje Turing. Úkolem moderátora je identifikovat výlučně na základě textů psaných strojem, textů, které jako jediné má k dispozici, povahu těchto hráčů. Musí proto klást otázky, které mu pomohou rozlišit mezi slovní promluvou a hmotnou, tělesnou realitou. Turing tvrdí, že pokud by moderátor nedokázal rozlišit výkon stroje v tomto testu od výkonu inteligentního, myslícího člověka, znamenalo by to, že stroje mohou myslet, že mohou být inteligentní.

V interpretacích *Turingova testu* bývá zdůrazňováno rozlišení mezi myslícím člověkem a myslícím strojem. Často se však zapomíná na první příklad, kterým Turing svou argumentaci otevírá – na rozlišení mezi mužem a ženou. Jestliže nemožnost rozlišit mezi člověkem a strojem potvrzuje, že stroje mohou myslet, co dokazuje první příklad, založený na nemožnosti rozlišit mezi mužem a ženou? Proč se zde u Turinga objevuje téma rodu? Proč A a B nejsou označeni jednoduše za lidi plnící rozdílné role v imitační hře? Jak souvisí argument založený na zrušení rodově specifických rozdílů ženských a mužských těl s úvahami o analogiích strojové a lidské inteligence? Proč je ten, kdo se snaží v *Turingově testu* hráči pomáhat, právě žena a proč ten, kdo mu chce škodit, je muž? Tohle nemůže být v článku autora, jakým byl Turing, náhoda! Vždyť v případě hráče C Turing ve svém článku jasně deklaruje, že na jeho pohlaví *nezáleží*. Tím nám však jakoby naznačoval (možná podvědomě?) že na pohlavích dalších hráčů *záleží*. Proč má počítač, má-li být inteligentní, plnit roli muže? Jaké podvědomé zkušenosti se ukrývají za takovýmto rozdělením rolí v imitační hře? Náhodně člověk nenapíše do filozofického článku muž nebo žena. Náhodně jim nepřidělí rozličné role... Natož to neudělá člověk tak přísně logicky myslící, jakým byl Turing.

Jak jsme se již zmínili, Hodges o Turingovi tvrdí, že izolované stroje, komunikující pouze pomocí dálkopisu, byly ideálem jeho vlastního života, ve kterém by nejraději komunikoval pouze prostřednictvím racionálních argumentů. Rod byl podle Hodgese ve skutečnosti falešnou stopou a jednou z mála pasáží Turingova článku, která nebyla napsána dokonale srozumitelně. Úspěšná imitace odpovědi ženy a muže by totiž nepotvrdila nic! Turing však nikde neuvádí, že první argumentace byla míněna jako protikladný příklad, naopak, tyto

dva příklady rétoricky spojuje jako paralelní. Rodový test a test člověk-stroj jsou určeny k tomu, aby potvrdily stejnou věc. Je možné, aby u geniálního matematika a vynikajícího rétora šlo skutečně pouze o špatnou argumentaci? Neschopnost vyjádřit myšlenou protikladnost mezi konstrukcí rodu a konstrukcí myšlení?

Hodges v již zmiňované biografii tvrdí, že Turing s oblibou zacházel se světem jakoby byl formální skládačkou. Do velké míry nebyl schopen rozlišit mezi tím, co se říká a tím, co se dělá, mezi promluvou a jednáním. Turing principiálně nerozuměl skutečnosti, že věci jako pohlaví, společnost, politika nebo tajemství, mohou být příkladem toho, jak to, co může být vysloveno, je omezeno ne inteligencí, ale restrikcemi v podobě imperativů toho, co má být uděláno. Můžeme pouze spekulovat o tom, proč Turing zahrnul téma rodu do testu, který měl prokázat možnost beztělesné či strojové inteligence. Skutečně nedokázal pochopit, že verbální performance rodu nemůže být srovnáváno s tělesnou zkušeností?

5 Magický trik zmizení těla

V roce 1952 byl Alan Turing zatčen, usvědčen z homosexuálního vztahu a následně souzen pro spáchání trestného činu nepřístojné nemravnosti mezi dospělými muži. Aby se vyhnul vězení, přijal alternativní trest v podobě nařízené roční hormonální léčby injekcemi estrogenu, které měly snížit jeho pohlavní touhu. Léčba měla však řadu nepřijemných vedlejších účinků, včetně růstu ženských prsníků. Jiným vedlejším účinkem jeho odsouzení bylo vyloučení z vládních výzkumných projektů pro ztrátu důvěry a bezúhonnosti. V roce 1954 spáchal Alan Turing sebevraždu.

Tajemstvím opředený Turingův osobní život použijeme jako klíč k pochopení konceptu "beztělesného" bytí, který se šířil kybernetickým diskurzem a dalšími oblastmi naší kultury a společnosti, od vědních disciplín, které se z něj vyvinuly (oblasti výzkumu umělé inteligence, umělého života, pokročilá robotika), přes umění (kybernetické umění a umění nových médií) až po kyberpunkové science-fiction, kde našel svou popularizovanou, zlidovělou podobu.

Tendence k zbavení se těla, která se pojí s užíváním digitálních technologií od jejich počátků, může na první pohled vést k domněnce, že souvisí s vnímáním světa, založeném spíše na racionálním a logickém úsudku, než na emocionálním a intuitivním prožitku; s popřením tělesné stránky lidské bytosti a s abstrakcí logické matematiky, spíše než s emocionálním výrazem a expresí fyzického gesta. Co když je ale všechno jinak a za mizením těla v digitálním obraze, kterého jsme svědky ve druhé polovině 20. století, se neskryvá převaha pragmatic-

kého myšlení lyotardovské techno-vědy v západní společnosti, nýbrž v tom, čehož je příkladem lidský osud Alana Turinga, fatálně spojený s tělesností? Vliv jeho myšlení na kybernetiku, oblast umělé inteligence, umělého života a informatiky obecně, byl tak velký možná i proto, že jeho osobní tragédie nebyla nikdy manifestována otevřeně. O to mocněji působila skrytě, sublimována v argumentaci *Turingova testu*. Interpretována jako logické selhání byla však v parafrázích a odkazech na Turingovo dílo postupně vymazána jako nadbytečná...

Turingův test můžeme chápat jako magický trik. Jako všechny dobré triky, je i tento test založen v první fázi na akceptaci předpokladů, které budou určovat, jak budeme interpretovat to, co uvidíme později. Předpokládejme, že klíčovým momentem argumentace Turinga není pasáž, ve které se snažíme určit kdo je muž, žena nebo stroj. Rozhodující moment – magický trik zmizení těla – přichází mnohem dříve – v okamžiku, kdy jsou hráči usazeni za dveře místnosti a jejich vůle, touhy a vnímání jsou vpleteny do imitační hry omezené na symbolické reprezentace znaků písma (dálnopisu, psacího stroje). Přistoupíme-li na tato pravidla hry, jsme chyceni ve smyčce kybernetického obvodu a bez odporu přijmeme různá strojová rozhraní, maskující odloučení jednajících těl od symbolických reprezentací jejich jednání.

Právě zde, v argumentaci *Turingova testu*, který nastavil program výzkumu umělé inteligence pro následující tři desetiletí, se nepozorovaně ustavuje kybernetický, digitální, informační či počítačový věk. Ve snaze zkonstruovat stroje, které mohou myslet, výzkumníci znovu a znovu uskutečňovali akt odstranění těla, který je však z určitého – třeba kulturologického – hlediska podstatou *Turingova testu*. Lidské tělo mizí v argumentaci kybernetického diskurzu, argumentaci směřující k analogickému chápání lidského a strojového, k přijetí konstruktu *kyborga*.

Z takového pohledu pozdější Turingovo zapletení s policií a soudním systémem v otázce homosexuality, představuje jakýsi epilog či anti-tezi *Turingova testu*. Jeho odsouzení a soudem nařízená hormonální léčba jeho homosexuality tragicky demonstruje exemplární ponaučení autora *Turingova testu*. Ponaučení o podstatném rozdílu mezi tvrzením a jednáním, jak se uplatňuje v nátlakovém řádu homofobní společnosti, z pozice moci uplatňující svoji vůli na tělech svých občanů.

I matematici, logici a informatici však mají svá těla. Těla, která jsou zásobárnou našich zkušeností, naší historií a identitou. Těla, na jejichž existenci a touhu upozorňuje již starozákonné Evino jablko. Bez těl by nebylo asi ani abstrakcí, ani symbolických světů, do kterých

se matematici tak rádi noří, a které informatici s takovým zanícením vytvářejí až do dokonalosti virtuálních světů a do nich vnořených virtuálních těl. Těla mají i robotické systémy, jejichž konstrukce nakonec přivedla alespoň k technokratickému poznání důležitosti těl pro všechno ostatní, co s člověkem obvykle (a s robotickými systémy prozatím méně obvykle) spojujeme¹⁰.

Tělo Alana Turinga, podrobené hormonální kastraci, ztrácející své původní rysy pod tukovými záhyby a s rostoucími, zaoblujícími se ženskými nadry, se samo stává argumentem, který zpochybňuje magickou moc triku *Turingova testu*.

6 Turingovo jablko

Turing byl/se za svůj omyl tvrdě potrestán/potrestal. Zemřel dne 7. června 1954, za zvláště inscenovaných okolností ve svém domě, na otravu kyanidem, kterým bylo impregnováno jablko, do kterého se zakousl. Neobvyklé okolnosti jeho smrti vedly k dohadům, jestli šlo o nehodu, sebevraždu nebo dokonce o vraždu.

Je třeba poznamenat, že nehoda není zcela vyloučená. Turing se v závěrečné části svého života intenzivně zabýval chemickými pokusy. Zabýval se hledáním podstaty biologické morfogeneze a tento zájem ho přivedl i k experimentálnímu zkoumání samovolného vytváření se rozmanitých obrazců v chemických roztocích v průběhu různých chemických reakcí (Turing, 1952). Záměrně tyto reakce vyvolával a se zájmem pozoroval jejich průběh. Pokoušel se s použitím výpočetní techniky té doby i automaticky generovat řešení pro některé problémy spojené s difúzními chemickými reakcemi. Ve strukturách rostlin předpokládal a nacházel struktury identické z matematického hlediska s *Fibonacciho posloupností* a jev, který takto objevil, nazval *Fibonacciho filotaxie*. Napsal o tomto jevu článek, který však zůstal za jeho života nepublikován (Turing, 1992). Tragická náhoda v laboratoři ne příliš pořádkumilovného experimentátora nemůže být tedy zcela vyloučena...

Matka možná neznala svého syna dobře. I proto byla ochotna připsat jeho smrt nešťastné náhodě, tragickému následku ochutnání díky nedbalosti otráveného jablka. Jeho přátelé se však domnívali, že zvolil tuto nejasnou formu sebevraždy záměrně, aby ji psychicky ulehčil a dal jí věrohodný argument, který by zastřel záměrnost jeho smrti. Spekulovalo se rovněž o vraždě. Ani tato hypoteze není zcela vyloučitelná, zejména kvůli tomu, že Turing byl minimálně v období 2. světové války intenzivně zapojen do činnosti tajných služeb a odhalení

¹⁰ Detailněji o roli těla pro robotické systémy pojednává např. (Brooks, 2002).

jeho homosexuality a následující skandál mohly tudíž představovat určité bezpečnostní riziko...

Sledujme Turingův konec a pak zmatené úvahy jeho matky, jak si je představuje a pak v závěru 17. a v 18. scéně své hry líčí Valeria Patera:

Alan: Cibule... Ano, chtěl jsem oddělit vrstvy cibule mysli, ale ve středu nic nebylo. *(Pauza)*

Doufal jsem, přál jsem si to najít.

(Vezme si jablko a zakousne se)

Jádro.

Ryzí život. *(Pauza)*

(Sedne si ke stolu, na kterém jsou ampule a laboratorní nástroje. Zatímco mluví, vezme krabičku s kyanidem, otevře ji, nasype obsah do malé nádoby a pomalu ponoří jablko do jedu.)

Počítač umí pouze otevřít okna logiky, kterými sám život unikne.

Iracionální a nenapodobitelný. *(Pauza)*

Počítač nikdy neocení pohádku jako malý chlapec. *(Pauza)*

Miluji Sněhurčino jablko. *(Pauza)*

Jeho odraz v zrcadle. *(Pauza)*

Kterým jsem prošel. *(Pauza)*

Imitační hru. *(Pauza)*

(Popěvujíc si uléhá na postel)

Jablko nořím do jedu

Na svět spavou smrt vedu...

7 Epilog

Vstoupí Ethel Turing, Turingova matka; mluví tiše, vede vnitřní spor o Alanově sebevraždě.

Ethel: Ten zvyk nemýt si ruce po experimentech ho zabil.

Experimentoval s kyanidem. *(Pauza)*

Byl posedlý představou jedu za svými nehty; nějaký se mu tam dostal. *(Pauza)*

Nedával dost pozor; pořádně si je nevydrhnul, a kdo ví... *(Pauza)*

Nikdy neuvěřím, že můj Alan si vzal život.

Byl zvláštní, jistě, ale ne zas natolik zvláštní. *(Pauza)*

Už byl zase klidný a skoro překonal trauma z té těžké léčby.

Ta ošklivá aféra nás vlastně sblížila. *(Pauza)*

A potom se znovu vrátil ke svým studiím počítačů, byl v Řecku... organizoval další cestu. *(Pauza)*

Ta myšlenka o tajné službě je jako něco mimo film. *(Pauza)*

Ano, přemýšlela jsem o tom, o všech těch tajných informacích, ale já jsem nikdy nic nevěděla. Co můžu říct? (*Pauza*)

Jeho absolutní nezájem o své tělo, o péči o vzhled, o mytí rukou... (*Pauza*)

Na střední škole byl úplně stejný...tvé ruce a nehty jsou špinavé, Alane, máš inkoust na límečku. (*Pauza*)

Umyj si ruce, Alane.

K o n e c

Literatura

Brooks, R. A.: *Flesh and Machines*. Pantheon Books, New York, 2002.

Gál, E., Kelemen, J. (sest.): *Mysel/Telo/Stroj*, Bradlo, Bratislava, 1992.

Gödel, K.: Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme. *Monatshefte für Mathematik und Physik* **38** (1931) 34-7.

Hodges, E.: *Alan Turing – The Enigma of Intelligence*. Unwin, London, 1983.

Church, A.: An unsolvable problem of elementary number theory. *American Journal of Mathematics* **58** (1936) 345-363.

Kelemen, J. a kol.: *Pozvanie do znalostnej spoločnosti*. Iura Edition, Bratislava, 2007.

Leawitt, D.: *Muž, který věděl příliš mnoho*. Dokořán, Praha, 2007.

McCulloch, W., Pitts, W. A.: A logical calculus of the ideas immanent in nervous activity. *Bulletin Math. Biophysics* **5** (1943) 115-133.

Patera, V.: Alan's Apple – Hacking the Turing Test. In: *Alan Turing – Life and Legacy* (C. Teuscher, Ed.). Springer, Berlin, 2004, pp. 9-41.

Plant, S.: *Zeros and Ones – Digital Women and New Technoculture*. Doubleday, London, 1997.

Turing, A. M.: On computable numbers, with an application to the Entscheidungsproblem. *Proc. London Mathematical Society* **42** (1936) 230-265; corrections **43** (1937) 544-546.

Turing, M. A.: Computing machinery and intelligence. *Mind* **59** (1950) 433-460 .

Turing, M. A.: The chemical basis of morphogenesis. *Philosophical Transactions of the Royal Society of London* **B237** (1952) 37-72.

Turing, A. M.: The morphogen theory of phyllotaxis. In: *Collected Works of A. M. Turing* (P. T. Saunders, Ed.). North-Holland, Amsterdam, 1992.

Umenie a emócie

Juraj Hvorecký¹

Abstrakt. Emocionálne pôsobenie umenia sa považuje za triviálny fakt. Nie je však jasné, ako môže literárne dielo, maľba alebo hudba emocionálne pôsobiť. Najprv ukážeme, ktorá časť umeleckého aktu je primárnou nositeľkou emócií. Potom upriamime pozornosť na existenciu dvoch súbežných stratégií vyvolávania afektívnych odpovedí. V závere naznačujeme, ako podvojná stratégia emocionálneho pôsobenia umenia podporuje zmiešaný pohľad na kategorizáciu emócií.

1 Na úvod o umení

Vzájomné spojenia medzi umením a emóciami sú dostatočne známe a rozmanité. Dokonca natoľko rozmanité, že by to mohlo zvädzať k presvedčeniu o nemožnosti povedať k tejto téme nič nové na tak malom priestore. Pretože sa zameriam na jeden špecifický problém prepojenosti týchto dvoch domén ľudskej existencie, dúfam, že čitateľ nesklamem a tému aspoň čiastočne osvetlím. Chcel by som však hneď na začiatku upozorniť, že moja úvaha sa nebude týkať ani oblasti estetiky a ani experimentálnych výsledkov kognitívnej vedy v oblasti emočnej percepcie umeleckej tvorby. Moja drzá odvaha pustiť sa trochu neobratne na široké pole vzťahov medzi ľudskou emocionalitou a umením bude trochu tlmená sústredením sa na oblasť, v ktorej sa cítim ako tak kompetentný. Inak povedané, budem špekulovať o tom, ako môže komplexný vzťah medzi emocionalitou a umením pomôcť filozofii mysle a tiež ako filozofia mysle vníma tento vzťah.

Na úvod niekoľko predbežných vymedzení tém, o ktorých nechcem písať. Po prvé, nebudem vonkoncom premýšľať nad tým, čo je umenie. Definovať umenie je zrejme ešte ťažšie ako definovať čokoľvek iné. Keďže sa v dejinách ukazuje, že sme príliš neuspeli s oveľa jednoduchšími kandidátmi na definíciu, rád ponechám túto medzeru nezaplnenú. Umenie budem jednoducho predpokladať a keďže ma nebudú zaujímať ani konkrétne diela, nechám otázku podstaty umenia nezodpovedanú. Nespokojencov, ktorí nástoja na tom, že bez predbežnej definície sa o nijakej téme nedá diskutovať, hádam zabrzdí sociologická definícia. Podľa nej je umenie civilizačným snažením, ktoré končí v galériách, koncertných sieňach, na policiach a uliciach. Odzbrojujúca definícia nemá nijaké aspirácie byť konečnou odpoveďou, pretože jej samozrejme chýba normatívny aspekt, ktorý je umeniu vlastný. Napriek tomu nebudem pracovať na jej zdokona-

¹ Filozofický ústav AV ČR, Jilská 1, 110 00 Praha 1, ČR, e-mail: juraj@hvorecky.com

lení a uspokojím sa s presvedčením, že všetci akosi rozumieme čo je umenie. Toto presvedčenie nenastoluje situáciu epistemologickej privilegovanosti a jasnosti, ale úplne postačí k nastoleniu základného argumentu tejto úvahy.

Pre mnohých prekvapivejšie bude moje dištancovanie sa od ďalšej dôležitej témy na pomedzí umenia a emócií. Nemám nijaké aspirácie riešiť problém miesta emócií v umení. Neviem, či je emocionálne pôsobenie pre umenie nevyhnutné, alebo len dôležité. Aj keď sa mi zdá, že všetky dôležité umelecké diela vyvolávajú silnú emocionálnu odozvu, predsa len nechcem úplne zavrieť dvere pred umením bez emócií. Trochu povrchno a možno provokatívne si dovoľím odhadnúť, že formálne vyšperkovaná umelecká tvorba, ako napríklad vizuálne geometrizmy alebo počítačná poézia, sú tvorené (a percipované) s čisto racionálnym úmyslom. Hoci emočne možno interpretovať každý objekt, intencia autora istým spôsobom vymedzuje interpretačné možnosti recipienta a tak si trúfam tvrdiť, že niektoré výsledky umeleckej tvorby nemajú vyvolávať (a nevyvolávajú) emocionálne reakcie. Pretože ide z mojej strany o nerefektívne tvrdenia bez potrebnej argumentácie, neprekvapia ma potenciálne odmietavé postoje čitateľov. Len zopakujem, že nech už je spojenie umenia a emócií akékoľvek, mnou skúmanej témy sa dotýka len okrajovo. Diskutovaný problém sa objaví, aj keď je emocionalita v umení okrajovou a náhodnou záležitosťou (čo sa mi, opakujem, zdá krajne nepravdepodobné).

2 Niečo málo o emocionalite

Už je načase, aby som prešiel k jadrú môjho skúmania. Bude ma zaujímať nasledovný problém: ak umenie emocionálne pôsobí, *ako to robí*? Ešte pred bližším priblížením významu tejto otázky chcem prípadného nespokojného čitateľa upozorniť, že práve podmienková formulácia problému mi umožňuje vyhnúť sa účasti vo vyššie spomínaných všeobecnejších diskusiách o mieste emócií v umení. Jednoducho predpokladám, že istá časť umeleckej tvorby je nabitá emóciami a nemusím sa púšťať do neľahkej odpovede na otázku po univerzálnosti tohto javu.

Pristavme sa najprv pri podotázke, čo to znamená, že niečo na nás emočne pôsobí. Emócie sú integrálnou súčasťou našej psychiky, fungujú komplexne a dozaista neplatí klasická predstava, že by boli v protiklade s racionalitou. Budem predpokladať štandardný prístup, ktorý rozlišuje tri komponenty každej emócie: fenomenálny, kognitívny a telesný. Fenomenálny komponent odkazuje na pocit, s ktorou je každá emócia spätá. Ak ma prekvapí milý darček, som potešený a môj pocit

sa vyznačuje špecifickou charakteristikou, ktorá ho odlišuje od ostatných mne známych pocitov. Je iný než pocity strachu, prekvapenia alebo zhnusenía. Pretože ide o privátny pocit mňa ako prvej osoby, periodicky sa objavujú skeptické argumenty o jeho realnosti. Pretože na tom náš argument nestojí, nebudeme ich tentoraz brať vážne.

Kognitívny komponent dopĺňa fenomenalitu emócie. K prchavosti a bezprostrednosti fenomenálneho aspektu emočného stavu pridáva širšie súvislosti a umiestňuje emóciu do globálneho rámca mysle. Pod tým predstavujeme, že emócia funguje vždy v interakcii s ďalšími mentálnymi stavmi a tie svojim pôsobením mení a usmerňuje. Naopak, ovplyvnené stavy zasa spätne špecifikujú a individualizujú danú emóciu. Funkcionalisticky povedané, každá emócia zohráva istú kauzálnu rolu, bez ktorej by jej existencia nebola uchopiteľná. Ak som vystrašený, lebo v hustom lese začujem praskot konárov, cítim sa nezávídeniahodne, no charakteristický pocit strachu nie je to jediné, čo vymedzuje môj konkrétny mentálny stav. Ľutujem, že som sa sem vôbec vybral, okamžite prehodnocujem svoju situáciu, na mieste zvažujem či je lepšie utekať alebo stáť bez pohnutia - to všetko sú sprievodné javy strachu a bez nich by strach nebol strachom. Fenomenálny a kognitívny obsah kráčajú ruka v ruke. Nestane sa nám, že by sme pocítovali nejakú emóciu a zároveň nemali nutkanie nejako na ňu odpovedať, podržať si ju, alebo ju naopak odvrhnúť.

Príklad s praskajúcim vetvami v lese zreteľne poukazuje na poslednú komponentu prežívania emócií – na ich telesnosť. Strach nielen silno pocíťujeme a rozličnými spôsobmi kognitívne spracovávame, ale zároveň telesne prežívame. Prudký adrenalinový šok, búšenie srdca alebo chvenie sú typickými znakmi veľkého strachu. Väzba s telom je natoľko silná, že existuje slávna tradícia v línii W. Jamesa, ktorá považuje telesné zmeny za štartér emočných stavov. James a jeho nasledovníci tvrdia, že telesná zmena predchádza emóciu a kauzálnu ju vyvoláva.

Pokiaľ ide o všeobecnú štruktúru emócií, pre naše potreby postačí rozčlenenie na dve časti, typické pre všetky intencionálne stavy. Emóciu charakterizujú objekt a mód vzťahovania sa k tomuto objektu. Vo vyššie uvedenom príklade z hustého lesa je objektom buď priamo silný zvuk, ale častejšie asi tvor, ktorého si predstavím ako jeho pôvodcu. V každom prípade sa bojím *niečoho*. Už zo spôsobu opisu plynie, že k takto vyčlenenej entite sa špecificky vzťahujeme. Nevisí osamotene vo vzduchoprázdne mysle, ale pútame ju k sebe práve módom strachu. Dokážeme si poľahky predstaviť, že v iných podmienkach by sme sa k tomu istému objektu vzťahovali iným módom, povedzme ak by sa identický zvuk alebo jeho údajný pôvodca objavi-

li v komediálnom filme. Domnievam sa, že z uvedených skutočností vyplýva pomerne zaujímavý predbežný záver. Súčasné začlenenie emócií medzi ostatné mentálne stavy je nedostatočné. S ohľadom na intencionálnu povahu emočných stavov sa mnohí autori klonia k ich zaradeniu medzi propozíčné postoje (jednoducho povedané, patria niekam medzi presvedčenia a túžby). Lenže neoddeliteľný fenomenálny komponent toto začlenenie spochybňuje. Preto iní autori radšej kladú emócie medzi pocity (k bolestiam, telesným pocitom a rôznym sensorickým javom). Kognitívny element je podľa nich dôsledkom dodatočných presvedčení, ktoré pocitovú zložku sprevádzajú. Nazdávam sa, že práve komplexnosť emócií bráni ich zatriedeniu do jednej z oboch skupín a je legitímne ich považovať za samostatnú kategóriu. Uvidíme, že skúmanie vzťahu emocionality a umenia podporí špecifické miesto pre afektívne javy a vyčlení ich voči ostatným dôsledkom konfrontácie s umením.

3 Spôsoby vplyvu

Komponenty emócií už poznáme. Aby sme mohli preskúmať emočné vplyvy umenia, musíme najprv zistiť, ktorú súčasť umeleckého aktu možno považovať za základného nositeľa emocionality. Termín *akt* využívam úmyselne, pretože za primárneho nositeľa v tomto prípade môžeme považovať aj entity mimo samotné dielo. Umelecký akt budem chápať ako triádu autor – dielo – recipient. Predpokladám, že primárnym nositeľom emócie môže byť iba jeden z trojice súčastí umeleckého aktu. V každej z troch súčastí nájdeme dôvody prečo práve ju považovať za nositeľa emocionality. Autor diela obvykle nejaké emócie pri tvorbe má a je dokonca možné, že ich umelecky stvárnil, prípadne tvorí iba preto, lebo má isté emócie. Dielo zasa prináša umelecké poslanstvo. Do neho autor vkladá konkrétne obsahy a z neho recipient čerpá umelecké dojmy. Zdá sa celkom samozrejmé, že emocionalita by sa mala primárne vyskytovať v samotnom diele. Poslednou súčasťou triády umeleckého aktu je recipient a aj u neho nájdeme dostatok dôvodov považovať ho za základ úvah o emocionalite umenia. Veď práve na neho dielo pôsobí a od neho pochádzajú svedectvá o afektívnom pôsobení konkrétneho artefaktu. Nebyť recipienta, dielo by nikdy emočne nepôsobilo.

Keďže všetky komponenty umeleckého aktu si nárokujú na emocionálnu jedinečnosť, musíme sa na problém pozrieť detailnejšie. Začnime u autora. Mnohé psychologicky orientované estetiky považujú emočnú náladu autora za základ tvorby. Umenie má odzrkadľovať autorovo vnútro, prípadne je výsledkom sublimácie nedoriešených psychických konfliktov (Freud). Ak by sme aj pripustili pravdivosť týchto

(pomerne pochybných) tvrdení, sústredenie sa na autora náš problém nerieši. Nejde nám o to, ako sa emócie do diela *dostali*, ale ako pôsobia. Keďže nás zaujíma pôsobenie emócií v umení, musel by byť náš záujem o autora zdôvodnený dodatočným argumentom, ktorý preukáže totožnosť medzi emocionalitou diela a autora. Lenže nijaký takýto argument neexistuje. Kompatibilita medzi emočným nábojom diela a jeho tvorcu je náhodná a ničím nezaručená. Je veľmi dobre možné, že autor postupoval úplne racionálne a dielo plné emócií je chladným kalkulom. Napokon, červená knižnica, hoci nie je umením, o tomto jave poskytuje dostatok dôkazov. Nie je dôvod veriť, že by to bolo v umení inak. Ak by sme sa aj dostali k dôkazom, ktoré by svedčili o mentálnom naladení autora v okamihu tvorby (denníkové záznamy, svedectvá), nebolo by jasné, ako ich interpretovať. Ak sa budú zhodovať s vyznením diela, bude to pre psychológistu skvelá správa. Ale ak budú v ostrom protiklade, len málokoho to prinúti zmeniť pohľad na emočné naladenie samotného diela. Autobiografické informácie sú obvykle veľmi zaujímavé, ale vypovedajú o tvorcovi, nie o jeho diele. Toto konštatovanie nijako neprotirečí bežnému chápaniu autora ako pôvodcu diela a tým pádom aj pôvodcu emócií v diele. V diele je práve to, čo tam autor vložil. To ale vonkoncom neznamená, že do diela autor vložil sám seba.

Analogická skepsa nás núti odmietnuť primárnosť emócií aj u recipienta, Hoci tu nenarážame na tak neprekonateľné prekážky ako je dávno zosnulý alebo neznámy autor, aj tu je korešpondencia medzi emočným vyznením diela a reakciou čitateľa náhodná. Ovplyvňujú ju vonkajšie okolnosti, predchádzajúce emocionálne naladenie a občas aj nepochopenie diela. Trochu ostro povedané, recipient môže diela prijať ľubovoľne. Takže aj keď dielo je nesamostatnou entitou, ktorej emočné vyznenie je priesečníkom intencií autora a schopností recipienta, predsa len netrpí idiosynkratizmom recipientových mentálnych stavov. Za primárneho nositeľa emócií budeme teda považovať samotné dielo. Ak vám prekáža absencia ľudského elementu, dielo môžete nahradiť ideálnym čitateľom. Ideálny čitateľ je myšlienkový konštrukt človeka, ktorý pristupuje k dielu s najlepšou možnou prípravou a absentujú u neho všetky javy, ktoré by mohli interpretáciu diela poškodiť. Jeho emocionálne reakcie sú vyvolané výsostne konfrontáciou s umeleckým dielom a plne odzrkadľujú emočné obsahy diela. Odnáša si z diela presne to, čo si z neho má odniesť. Ideálny príjemca je umeleckým dielom transformovaný a berie si z diela autorom vložený obsah, vrátane príslušných emócií. Predstavuje imperatív nášho recipovania a normu správnosti dekodovania umeleckého odkazu diela.

4 Intencionálna štruktúra

Keďže sme našli miesto umeleckého aktu, ktoré je primárnym nositeľom emócií, môžeme konečne podrobnejšie preskúmať metódy emočného pôsobenia umenia. Vzhľadom na bohatosť umeleckých foriem sa zdá, že musia existovať prinajmenšom dve takéto metódy. Jedna z nich využíva kognitívny komponent emócií, druhá sa spolieha na fenomenálny komponent. (Je možné, že virtuálna realita a iné dômyselné interaktívne podoby umenia dokážu vyvolať emócie aj pomocou telesného komponentu, ale to nie je predmetom tohto príspevku.) Paradigmatickým príkladom prvého typu je literatúra a klasické vizuálne umenie. Druhý typ objavíme v hudbe a zrejme aj v abstraktnejšom vizuálnom umení. Aj ďalšie druhy umenia, napríklad divadlo a film, spadajú pod oba vymedzené typy. V klasickej podobe spadajú obvykle pod kognitívnu metódu, v neštandardných abstraktnejších občas aj pod fenomenálnu.

Pozrime sa najprv na kognitívnu metódu ukotvenia emócií v umeleckom diele. Už sme upozornili na dvojdielnu štruktúru emócií, v ktorej je emočný mód namierený na nejaký objekt. Pozoruhodnou súčasťou takéhoto intencionálneho vzťahu je transparentnosť objektu. Objekt sa do mentálneho stavu dostáva priamo, nesprostredkovane. Keď sa zľaknem psa, môj strach sa týka konkrétneho psa a ničoho iného. Ten istý pes môže figurovať v iných mentálnych stavov a zároveň v danej situácii je dostatočným dôvodom môjho strachu. Intencionálna štruktúra zaisťuje spätosť emócií s ich objektmi. Práve tento jav využívajú literáti a vizuálni umelci. Ak pred nás postavia v istom kontexte istý objekt, emócia sa dostaví. Je dôležité upozorniť, že kontext samotný nie je ničím iným než ďalším objektom.

Transparentnosť si nemýľme s reprezentačnými vlastnosťami objektu. Nejde o to, že literárny príbeh alebo obraz k niečomu odkazuje. Reprezentačné schopnosti sú vo väčšine diel neodškriepiteľné. Transparentnosť ale nie je vzťahom medzi dielom a tým, čo zastupuje, ale medzi mentálnou udalosťou a jej objektom. Je úplne jedno, či ide o objekt zo sveta umenia, alebo z reálneho sveta. Preto transparentnosť nie je to isté ako reprezentačná vlastnosť, aj keď na mentálnej úrovni sa môžu prekrývať. Dielo nemusí odkazovať na svet, nemusí nás nútiť aplikovať emočné stratégie z reálneho sveta na svet umelecký a predsa je transparentné. Jeho transparentnosť je postavená na skutočnosti, že obsahuje objekty, ktoré si vyžadujú aplikáciu príslušného emočného módu. To znie možno triviálne, ale iba do okamihu, keď si uvedomíme dôsledok tohto kroku. Emočná sila diela nezávisí od opisu príslušného emočného módu. Smutnú scénu nevyvoláme tak, že opíšeme smútok všetkých zúčastnených, ale popíšeme dej, ktorý smútok vy-

voláva. Vhodne popísaná scéna sa o vyvolanie smútku postará sama. Emócie stoja na svojich objektoch, nie na snahe navodiť ich čo najpodrobnejšou deskripciou emocionálneho módu. Silný moment nestojí na detailoch telesných a vnútorných zmien, sprevádzajúcich emóciu, ale na vhodnom vykreslení podmienok, ktoré viedli k silnej reakcii. Presný opis módu nie je samoúčelný. Môže nás poučiť o aspektoch emocionality, o ktorých sme neuvažovali, prípadne unikali našej pozornosti. Lenže opisy módov neprispievajú k emocionálnemu vyzneniu umeleckého diela. Dôležitejšia je transparentnosť opisu, pretože tá si korešpondujúci mód nájde aj sama. To má dôsledky pre autorov emočne nabitých diel. Úlohou autora nie je podrobne opisovať ako sa kto cíti, ale prispieť k transparentnosti obsahu. Ten nás potom sám naviguje pri afektívnej recepcii. Autor teda musí dielo vymodelovať do podoby, ktorá v nás vyvoláva emócie úplne rovnako ako kontexty a objekty v bežnom svete. Z uvedeného je jasné, prečo zaradujem literatúru a vizuálne umenie pod kognitívne cesty k emocionalite. Keď umelec predostrie recipientovi rozvrh svojho sveta, zároveň ho žiada, aby sa v novom svete sám zorientoval. Čitateľ alebo divák musí na dielo všemožne reagovať, očakáva sa od neho globálna odpoveď. Jej prirodzenou súčasťou je emocionálna odpoveď. Odpoveď ideálneho recipienta je presne v intenciách autora a jeho diela, my ostatní odpovedáme podľa našich obmedzených interpretačných schopností. No ideálny a ani bežný čitateľ či divák nereagujú emočne na mód, ale na objekt diela.

5 Fenomenalita

Intencionálna štruktúra je však typická len určité druhy umenia. Rozvrh deja alebo scény sa v literatúre deje slovami a vo vizuálnom umení ťahmi štetca, tvarovaním hliny a ďalšími nástrojmi. Intencionálny objekt emócie vieme v prípade potreby verbálne zachytiť aj bez emocionálnej reakcie. Nie je neobvyklé, že pokus prerozprávať cieľový objekt končí fiaskom, pretože synonymita nezaručuje intenzionálnu identitu a čaro diela sa úplne vytratí. To potvrdzuje dobre známu skutočnosť, že intencionálny objekt má často intenzionálne vlastnosti. Keď nie je skonštruovaný presne podľa intenzionálnej schémy, ktorá vyvoláva emócie, stráca všetku svoju silu. V každom prípade ale platí, že verné pretlmočenie emočný náboj zachováva.

Zamerajme sa teraz na druhy umenia, ktoré nijakú intencionálnu štruktúru nemajú, nedajú sa prerozprávať a napriek tomu emócie vyvolávajú. Ide o hudbu (najmä klasickej) a abstraktné vizuálne umenie. Hoci sa mnohé hudobné diela pýšia názvami, evokujúcimi potenciálne intencionálne objekty, je ilúziou im intencionálne obsahy

pripisovať. Nech to znie akokoľvek cynicky, Smetanova *Má vlast* o Čechách nie je. A nie je ani o nijakej inej krajine. Jednoducho nie je o ničom. Obsahy si môžeme do diela primýšľať, ale stratégia primýšľania je náhodná a neplynie z diela, ale z vedľajších okolností. V diele musí byť niečo iné, čo v nás vyvoláva emočné reakcie na tento druh tvorby.

Existuje viacero návrhov, čo stojí v pozadí emocionálnej sily tohto typu umenia. nemá zmysel podrobne sa nimi zapodievať. Detailné návrhy sú podľa mňa buď predčasné alebo pomýlené. Z filozofického hľadiska je najdôležitejšou nasledovná skutočnosť: nech už je metóda evokovania emócií týmito dielami akákoľvek, nestojí na intencionalite, a ani na intencionálnych objektoch. Je možné, že akékoľvek úvahy o emocionálnej sile hudby skončia pri diskusií o štrukturálnej *podobnosti* medzi hudbou a emóciou. Naznačujú to niektoré (triviálne) príklady. Napríklad smutná vrbá, ktorá v nijakom prípade nemôže byť smutná, sa spája s emocionálnou zrejme kvôli svojej zhrbenosti. Ide o podobnosť s depresívnymi stavmi, keď sa cítíme „na dne“, alebo ako hovoria Angličania, jednoducho „down“. Možno aj symfónie a piesne nesú znaky, ktoré ich pripodobňujú ľudským emocionálnym stavom. Odborná literatúra je plná produktívnych návrhov, ktoré aspekty hudby by mohli byť potenciálnymi nositeľmi podobností. Medzi kandidátov patrí rytmus, plynutie alebo výstavba diela. Obávam sa, že skutočne kvalifikovanej odpovede sa len tak nedočkáme. Analýzy podobnosti ukazujú, že ide o veľmi nejednoznačný fenomén aj v jednoduchších oblastiach než práve umenie a emócie.

Pretože nám nezáleží na detailoch vysvetlenia emocionálneho pôsobenia hudby, pozrime sa na problém z väčšieho odstupu. V prípade literárnej a vizuálnej tvorby sme konštatovali, že emocionalita diela sa objavuje pomocou zamerania sa na vykreslený objekt. Pri dekódovaní umeleckého diela sa spoliehame na kognitívne procesy, ktoré recipovaný objekt a jeho kontext spracujú emočných módom. Hudba a abstraktné umenie takéto uchopenie emocionality neumožňuje. Napriek tomu pri vnímaní týchto druhov umenia *pocitujeme* emocionálne zmeny. Emočné obsahy diel teda musia byť vyvolávané fenomenálne. Lenže tým sa od diel s jasným intencionálnym objektom nelíšia. Aj tie predsa vyvolávajú pocity ako neoddeliteľnú súčasť emočnej reakcie. Takže rozdiel nespočíva vo vyvolávaní pocitov, ale jeho etiológii. Kým v literatúre a ďalších intencionálnych módoch umenia sa pocity objavujú v súvislosti s kognitívnym spracovaním intencionálneho objektu, u hudby je prítomná len fenomenálna stránka emocionality. Intencionálne umenie vyvoláva dva aspekty emócií súčasne, fenomenálnemu umeniu postačí jeden aspekt.

Niekoľko by úvahy o podvojnóm spôsobe vyvolávania emócií mohli viesť k presvedčeniu o dominantnej úlohe fenomenality. Keďže v oboch druhoch umeleckého vyvolania emócií je prítomná fenomenalita a kognitívny mód iba raz, možno je kognitívny aspekt nadbytočný. Ide však o nesprávny záver. Intencionálnym objektom vyvolané emócie sa odlišujú od čisto fenomenálnych svojou silou a špecifickosťou. Sú namierené na konkrétny objekt a táto zacielenosť zvyšuje ich intenzitu. Strach z praskajúcich konárov v lese je podstatne silnejší než strach vyvolaný tmnou hudbou. Sila pocitu zrejme závisí od špecifickosti kontextu, ktorý v hudbe nevyhnutne absentuje.

Dôležitosť viac aspektovosti emócií plynie aj zo mimovoľného objavovania sa intencionálnych motívov pri recepcii neintencionálnych diel. Mimovoľné asociácie pocitov s objektami sú pri počúvaní hudby časté a zrejme odkazujú na východiskovú viac aspektovosť emócií. Nazdávam sa, že takéto asociácie sa objavujú bez ohľadu na prípadne intencionálne pomôcky, ktorými autor, kurátor alebo performer pripravujú recipienta na umelecký zážitok. Takže to nie sú názvy diela alebo rozbory kritikov, ktoré asociácie vyvolávajú, ale ideo automatický vedľajší efekt fenomenality. Mimovoľné naplnenie mysle obsahom, ktorý je nejakým späť s recipovanou fenomenalitou ukazuje, že spojenie medzi intencionálnymi a fenomenálnymi druhmi vyvolávania emócií v umení je veľmi pevné.

6 Záver

Existencia dvoch spôsobov vyvolávania emočných reakcií u recipientov umenia má zásadný význam pre debatu o mieste emócií v ľudskej mysli. Ukazuje sa, že na rozdiel od propozíčných postojov a čistých pocitov sú emócie oveľa štruktúrovanejšie a ich viac aspektovosť nie je náhodná. Nedá sa povedať, že by jadrom emócií bola ich intencionálna štruktúra, pretože hudba a abstraktné umelecké smery dokážu vyvolať afektívne reakcie bez prítomnosti emocionálneho objektu. Na druhej strane, bez intencionálneho objektu sú emócií nešpecifické a nevýrazné. Úvahy o emočnom pôsobení umenia teda posilňujú pozíciu, podľa ktorej nemožno redukovať emócie ani na propozočné postoje a ani na pocity. Znemožňuje to ich esenciálna viaczložkovosť.

Naše úvahy zároveň ukazujú, že umenie je dobrým testovacím okruhom pre teóriu emocionality. Kým v bežnom živote sa obvykle emócie objavujú v celej svojej zložitosti, umenie dokáže selektívne upriamiť svoje pôsobenie na jednotlivé ich aspekty. Preto sa filozofia mysle musí zaujímať o teóriu umenia. Opačne to ale neplatí. Hoci výsledky filozofického výskumu emócií môžu byť pre estetikov prí-

nosné, nekonštituujú test udržateľnosti estetických úvah. Emócie nie sú pre umenie nevyhnutné, ale umenie o emóciách vypovedá niečo veľmi dôležité.

PodĎakovanie: Tento príspevok vznikol za podpory Grantovej agentúry Akadémie vied ČR v rámci grantovej úlohy KJB900090802. Ďakujem poslucháčom Kognitívneho klubu na VŠE v Prahe za diskusiu základných motívov tejto práce počas mojej prezentácie v januári 2008.

Literatúra

- [1] Casey, J.: Emotion and Imagination, *Philosophical Quarterly*, **34**, (1984) 1-14.
- [2] Oxenhandler, N.: The Changing Concept of Literary Emotion: A Selective History, *New Literary History* **20** (1988) 105-121.
- [3] Stecker, R.: Expression of Emotion in (Some) Arts, *The Journal of Aesthetics and Art Criticism* **42** (1984) 409-418.
- [4] Stecker, R.: Expressiveness and Expression in Music and Poetry, *The Journal of Aesthetics and Art Criticism* **59** (2001) 85-96.

Modely gramatiky života

Alica Kelemenová^{1 2}

Abstrakt. Príspevok poskytuje prehľad gramatík a gramatických systémov, ktoré sa využívajú pri modelovaní niektorých aspektov a prejavov (umelého) života. Postupne predstavíme Lindenmayerove systémy, kolónie gramatík, ekogramatické systémy a membránové systémy.

1 Úvod

Spoznávanie biologickej reality a vytváranie modelov biologických procesov umožňujú okrem tradičných metód a štandardných biologických prostriedkov a postupov aj prostriedky a metódy deduktívnych vied, matematiky a teoretickej informatiky. Tieto disciplíny sa podieľajú na hlbšom a detailnejšom spoznávaní skúmaných objektov a disponujú v mnohých prípadoch prediktívnou schopnosťou.

Jedným z prostriedkov, ktoré v niekoľkých posledných desaťročiach prispievajú k štúdiu biologických javov, sú pojmy, metódy a prostriedky vyvinuté v rámci *teórie formálnych gramatík a jazykov*. Ide o teóriu, ktorá vznikla v druhej polovici minulého storočia, a ktorá zaznamenáva významný rozvoj hlavne vďaka jej súvislosti s výskumom možností a limitov využívania výpočtovej techniky. Na formalizáciu jej základných objektov boli využité štruktúry diskkrétnej matematiky umožňujúce štúdium vývoja a zmien štruktúr objektov, na základe výpočtov určených prepisovacími pravidlami a typom odvodení.

Formálne gramatiky pracujú s reťazcami symbolov, slovami, ktoré pretvárajú v diskrétnych intervaloch pomocou svojich pravidiel. Slová, ktoré je gramatika schopná vytvoriť určujú jej jazyk. Pravidlá gramatiky stanovujú možnosti, ako nahradiť v reťazci symbol alebo (pod)reťazec symbolov novým reťazcom. Týmto spôsobom poskytujú prostriedok určujúci zmenu či vývoj slova, reťazca, alebo všeobecnejšie štruktúry na základe povolených lokálnych zmien vnútri reťazca. Podstatnú úlohu v tomto procese prepisovania reťazcov hrá aj spôsob, akým sa pravidlá pri odvodzovaní používajú, teda skutočnosť, koľko pravidiel a na ktorých miestach reťazca je možné použiť pri jednom prepise. Teória poskytuje širokú škálu formálnych gramatík a nimi určených formálnych jazykov vhodných pre potreby mo-

¹ Ústav informatiky, Slezská univerzita v Opavě, Bezručovo nám. 13, 746 01 Opava, Česká republika, E-mail: kelemenova@fpf.slu.cz

² Katedra informatiky, Pedagogická fakulta, Katolícka univerzita, nám. A. Hlinku 56/1, 034 01 Ružomberok, Slovensko

delovanej biologickej reality, líšiacich sa spôsobom používania pravidiel pri transformácii reťazcov a to od typicky sekvenčných gramatík (prepisujúcich vybraným pravidlom jednu lokalitu v reťazci), cez gramatiky používajúce jedno pravidlo na simultánny prepis na viacerých miestach reťazca, až po Lindenmayerove systémy, ktoré prepisujú v každom kroku výpočtu súčasne všetky symboly reťazca (pozri napríklad [32].) Výber použitých pravidiel je v niektorých typoch gramatík závislý len od aktuálne prepisovaného reťazca, iné typy gramatík poskytujú možnosť obmedziť poradie aplikácie pravidiel ďalšími špecifikáciami.

2 Lindenmayerove systémy

Lindenmayerove systémy sú typickým a v súčasnosti detailne rozpracovaným predstaviteľom generatívneho modelu široko využívaného pre biologické ciele. Ich autora Aristida Lindenmayera viedla snaha vytvoriť taký model vývoja jednoduchých organizmov, ktorý by umožnil nielen opísať skúmanú biologickú predlohu – jednoduchý organizmus, ale aj vypovedať o jeho možnom vývoji, správaní sa, charakterizovať vlastnosti organizmu a biologicky relevantné skutočnosti. Okolnosti, ktoré predchádzali zavedeniu týchto vývojových systémov priblížil autor v rozhovore z roku 1983 ([27],[22]).

A. Lindenmayer, profesiou teoretický biológ, bol značne zbehlý aj v práci s formálnymi metódami matematiky a matematickej logiky, čo mu umožnilo aktívne sledovať dianie súvisiace s nástupom výpočtovej techniky. Inšpirovaný vtedajšou novinkou, konečnými automatmi a poznatkami o možnosti spájať ich do reťazcov, rozpoznal analógiu, ktorá ho priviedla k myšlienke využiť konečný automat ako model (určitých komunikačných aspektov) bunky. Pospájanie konečných automatov do reťazca mu umožnilo modelovať organizmy s jednoduchou vláknitou štruktúrou. Biologická predloha vyžadovala, aby všetky automaty v takýchto reťazcoch pracovali synchronne. Na rozdiel od technicky motivovaných modelov vytvorených prepojením konečných automatov, ktorých vlastnosti boli v tom čase už podrobne preštudované, a ktoré sa vyznačovali nemenným počtom komponentov, vyvstala pred Lindenmayerom nová skutočnosť. Bolo potrebné, aby sa jeho reťazce automatov mohli počas vývoja predlžovať alebo skracovať podľa toho, ako sa bunky v modelovanom organizme množia, či zanikajú. Tento nový rozmer sa ukázal byť podstatným a určujúcim nielen pre vznik a rozvoj teórie L systémov, ale v širšom meradle aj pre ďalší rozvoj a smerovanie samotnej teórie formálnych jazykov.

Informácie o pôvodnom modeli vývojových systémov [28] môže čitateľ nájsť aj v publikácii *Mysel, inteligencia a život* [21].

Dnes má tento model vývojového systému známy pod názvom L systém tvar paralelne pracujúcej formálnej gramatiky. Transformáciu pôvodného automatového modelu na systém paralelne pracujúcich prepisovacích pravidiel navrhol a úspešne zaviedol G. Rozenberg, dlhoročný spolupracovník A. Lindenmayera a jeho sused v holandskom Huist ter Heide.

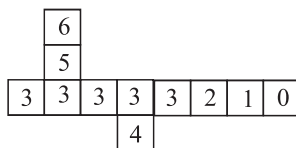
Uvedieme ukážku toho, ako generatívne pravidlá umožňujú modelovať vývoj rastliny (vetviacej sa štruktúry). Jednotlivé bunky rastliny označíme číslami 0 až 9, ktoré charakterizujú ich vnútorné stavy a určujú ich vývojový potenciál. Vývoj buniek určujú nasledujúce pravidlá

$$\begin{array}{cccccc} 0 \rightarrow 10 & 1 \rightarrow 32 & 2 \rightarrow 3(4) & 3 \rightarrow 3 & 4 \rightarrow 56 & (\rightarrow (\\ 5 \rightarrow 37 & 6 \rightarrow 58 & 7 \rightarrow 3(9) & 8 \rightarrow 50 & 9 \rightarrow 39 &) \rightarrow) \end{array}$$

Napríklad prvé pravidlo vyjadruje, že bunka 0 sa v ďalšom kroku odvodenia (v ďalšom vývojovom štádiu) rozmnoží a bude nahradená dvoma bunkami označenými 1 a 0. Všetky bunky tohto systému sa množia až na bunku 3, ktorá sa v priebehu vývoja nemení.

Z pôvodnej jednej bunky 0 systém postupne vytvára reťazce 0, 10, 3210, 33(4)3210, 33(56)33(4)3210, 33(3758)33(56)33(4)3210, ...

Reťazec v zátvorke môžeme interpretovať ako vetvu nad bunkou, ktorá sa pred príslušnou zátvorkou nachádza. To umožní graficky znázorniť napríklad reťazec 33(56)33(4)3210 schématickou vetvičkou na Obrázku 1. Ako pribúdajú v procese odvodenia ozátvorkované reťazce, tak budú pribúdať ďalšie vetvičky rastlinky. (Zátvorky sú používané v systéme z technických dôvodov a preto im náležia pravidlá uvedené v poslednom stĺpci, ktoré zachovávajú ich stabilitu.)



Obrázok 1.

L systémy môžu poskytnúť rozličné biologicky relevantné informácie. Uvedieme príklad systému schopného regenerovať svoj dospelý reťazec (reťazec, ktorý sa už v priebehu vývoja nemení) z akéhokoľvek jeho „zvyšku“. Pre tento účel použijeme interaktívny L systém, u ktorého určujú pravidlá prepis písmena v závislosti od obidvoch susedných písmen. Pravidlo zapísané v tvare $(a,b,c) \rightarrow u$ umožňuje nahradiť písmeno b v kontexte $\dots abc \dots$ reťazcom u . Pri prepise prvého alebo posledného písmena reťazca použijeme symbol p na označenie, že bunka susedí s prostredím.

Príklad L systému, ktorý dokáže pri poškodení regenerovať reťazec abc z akéhokoľvek svojho zachovaného podreťazca a vygenerovaný reťazec abc ďalej nemení, má nasledujúce pravidlá

1. $(p, a, p) \rightarrow ab$
2. $(p, b, p) \rightarrow c$
3. $(p, c, p) \rightarrow bc$
4. $(p, a, b) \rightarrow ab$
5. $(a, b, p) \rightarrow c$
6. $(p, b, c) \rightarrow ab$
7. $(b, c, p) \rightarrow c$
8. $(a, b, c) \rightarrow \varepsilon$

Ak zo slova abc zostalo len a alebo b , systém generuje reťazce nasledovne

a, ab, abc alebo b, c, bc, abc

Uvedené sekvencie odvodení udávajú tiež regeneráciu ostatných podreťazcov c, bc, ab .

Podrobnejšie informácie o výskume L systémov a literárnych zdrojoch z tejto oblasti poskytuje práca [21].



Obrázok 2. Tri knihy vydané vydavateľstvom SPRINGER, ktoré sú venované počítačovej grafike prírodných živých objektov.

L systémy sa stali predmetom zaslúženého záujmu odborníkov z počítačovej grafiky v súvislosti s grafickou reprezentáciou generovaných objektov, a dnes predstavujú významnú súčasťou tejto oblasti informatiky. Bohatý zdroj informácií o tomto smere výskumu poskytuje [www stránka http://algorithmicbotany.org](http://algorithmicbotany.org). Po úvodnej pôsobivej knižnej publikácii Algoritmická krása rastlín [31] úspešne prezentujúcej algoritmické postupy pri vytváraní štruktúr typických pre vývoj rastlín, sme svedkami spoznávania algoritmického správania ďalších graficky zaujímavých výtvorov prírody ako ilustrujú tituly Algoritmická krása morských rias, húb a koralov, Algoritmická krása mušlí [11], [29] a ďalšie. (Pozri obrázok. 2)

3 Ekogramatické systémy

L systémy sme doteraz prezentovali ako nástroj, ktorý umožňuje pracovať s vývojom jednej rastliny, jedného organizmu. Formalizmus L

systémov poskytuje vhodný a dostatočný priestor realizovať prirodzeným spôsobom aj delenie organizmov a zachytiť vývoj celej skupiny potomkov. Podobne ako sme v prípade vetvičky použili špeciálne symboly, zátvorky, na vyznačenie vetvenia organizmu, môžeme teraz použiť iný špeciálny symbol, spravidla \perp , na vyznačenie miesta, kde sa organizmus rozdelí. Symbol \perp sa priebehu vývoja nemení, preto jeho prepis určuje jediné pravidlo $\perp \rightarrow \perp$. Takto môže L systém určený pravidlami

$$1 \rightarrow 23, \quad 2 \rightarrow 1 \perp 3, \quad 3 \rightarrow 3, \quad \perp \rightarrow \perp.$$

vytvárať z reťazca 13 postupne reťazce

$13, 233, 1 \perp 333, 23 \perp 333, 1 \perp 33 \perp 333, 2 \perp 33 \perp 333, 1 \perp 3 \perp 33 \perp 333, \dots$

reprezentujúce stále narastajúcu populáciu organizmov. Správanie každého z takto generovaných organizmov určujú dané vývojové pravidlá. Novovytvorené organizmy sa rozvíjajú ako nezávislé objekty, bez vzájomného ovplyvňovania a bez ovplyvňovania prostredím, v ktorom sa nachádzajú.

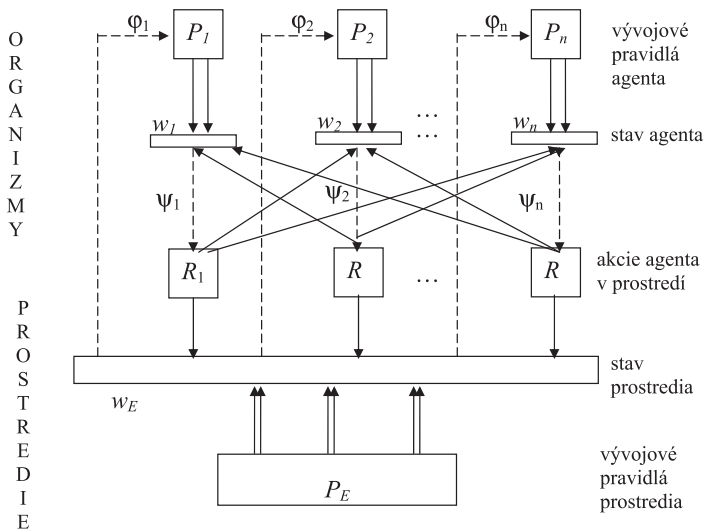
Zovšeobecnieme teraz sledovaný problém: Položme si otázku, či je možné charakterizovať pomocou (systému) gramatík aj vývoj ekosystému, tj. skupiny rôznych organizmov s ich vzájomnými väzbami a väzbami na prostredie? Tento problém sme začali riešiť v prácach [6] a [7].

Vychádzali sme z analýz na požiadavky modelu ekosystému, ktoré prezentovali Farmer a Bellin v práci [10] v tvare šiestich základných axiém. Mali sme pri tom k dispozícii techniku a poznatky rozvíjajúcej sa teórie gramatických systémov [5]. Naše snaženie vyústilo do modelu nazvaného ekogramatický systém, ktorý je schematicky znázornený na obrázku 3. Popíšeme komponenty tohto modelu a ich vzájomné väzby.

Ekogramatický systém je tvorený prostredím a niekoľkými organizmami. Stav prostredia w_E a stavy organizmov w_1, \dots, w_n reprezentujú reťazce. Individuálny vývoj každého z organizmov je určený jeho vlastným L systémom. Ten určuje vývoj organizmu v ideálnej situácii, keď ho v ekosystéme neovplyvňujú iné organizmy ani stav prostredia. Individuálny vývoj prostredia charakterizujeme tiež L systémom, ktorý určuje správanie prostredia v situácii, keď v prostredí nie sú organizmy.

Kľúčovým pre vytvorenie ekogramatického systému je spôsob, akým sú v modeli vymedzené vzájomné vplyvy jednotlivých zložiek ekogramatického systému. Na základe požiadaviek formulovaných

v práci [10] sme postupovali nasledovne: Pravidlá vývoja prostredia P_E ekogramatického systému sú fixné, systém nijakým spôsobom neumožňuje obmedziť či rozšíriť ich použitie. Stavom prostredia však môžeme dočasne obmedziť (zablokovať) použitie niektorých vývojových pravidiel organizmu. Na obrázku 3 tento vzťah vystihuje funkcia výberu φ . V súlade s biologickou predlohou sme umožnili organizmom ovplyvňovať sa vzájomne a vplývať tiež na vývoj prostredia prostredníctvom skupiny gramatických pravidiel R , určujúcich akcie organizmu v systéme. Organizmus vykoná v jednom okamihu najviac jednu z možných akcií na zvyšku ekosystému tým, že zasiahne do vývoja niektorého iného organizmu, alebo ovplyvní vývoj prostredia. Akcie organizmov sú „významnejšie a prieraznejšie“ ako vývoj prostredia, čo znamená, že pri synchronizovanej práci celého systému, prepíšu pravidlá z P_E , P_1 , P_2 , ..., P_n len tie časti reťazcov, ktoré v danom okamihu neovplyvnia akcie žiadneho z organizmov. Výber akčných pravidiel môže byť limitovaný dosiahnutým stupňom vývoja organizmu, čo v modeli umožňuje výberová funkcia ψ .



Obrázok 3

Takéto systémy umožňujú sledovať súčasne stav prostredia aj stavy jednotlivých organizmov, určiť *jazyk prostredia* ako množinu možných stavov prostredia a tiež vývojové reťazce a *jazyk jednotlivých organizmov*.

Príklady, ilustrujúce činnosť ekogramatických systémov, nájde čitateľ v práci [19]. Základné vlastnosti modelu prezentujeme na nasledujúcom jednoduchom príklade.

Prostredie je vytvorené reťazcami symbolov a, b, c a riadi sa vlastnými vývojovými pravidlami

$$a \rightarrow ab \quad b \rightarrow b \quad c \rightarrow cc.$$

Vývoj prostredia začínajúceho reťazcom aab je nasledovný:

$$aab, ababb, abbabbb, \dots$$

Každý z vytvorených reťazcov obsahuje dva výskyty písmena a a neobsahuje výskyt písmena c . Na každom kroku sa reťazec predlžuje dvoma symbolmi b , každý z nich je umiestnený bezprostredne za symbolom a . Za druhým výskytom a je vždy o jedno b viac ako za prvým výskytom a .

Organizmus: Vložme do tohto prostredia organizmus s jeho vnútorným správaním určeným symbolmi j, k a pravidlami $j \rightarrow k, k \rightarrow j$. Na prostredie organizmus vplyva akciou odstraňujúcou jeden výskyt a , t.j. pravidlom $a \rightarrow \varepsilon$.

Organizmus cyklicky strieda stavy určené reťazcami jk, kj . Prostredie sa vyvíja takto:

$$aab, abb, bb, bb, \dots$$

Vplyvom organizmu prostredie postupne stráca symboly a a tým aj schopnosť vytvárať nové výskyty symbolu b a degeneruje na reťazec bb .

Ďalší organizmus: Pridajme do prostredia ďalší organizmus s rovnakým vnútorným správaním ako má prvý organizmus a s akciou $b \rightarrow c$, ktorá zapríčini že bunka zmení svoj vnútorný stav z b na c .

Vývoj prostredia s obidvoma organizmami je takýto:

$$aab, abc, ccc, ccccc, ccccccccc, \dots$$

Prostredie vplyvom druhého agenta mení kvalitu. Táto je reprezentovaná bunkami v stave c , ktoré zapríčínajú exponenciálny rast prostredia.

Príklad ilustruje dramatické dôsledky, ktoré môže mať činnosť organizmu na vývoj prostredia. Malá *lokálna* zmena spôsobená organizmom (zánik *jednej* bunky, zmena kvality *jednej* bunky) môže viesť k extrému, kedy sa pôvodne pozvoľný, lineárny nárast prostredia zastaví a udržuje na "minimálnej hodnote", i k extrému, charakterizujúcu exponenciálne zväčšenie prostredia.

V poslednom období sa venuje pozornosť skúmaniu niektorých špeciálnych typov ekogramatických systémov. Sú to napríklad mo-

nokultúry, systémy s identickými organizmami. Ukazuje sa, že účinnosť týchto systémov sa len málo líši od účinnosti všeobecných ekogramatických systémov [23], [1].

S modelmi ekogramatických systémov s pasívnym prostredím sa stretávame v prácach [24], [33]. Modely ekogramatických systémov s lokalizáciou organizmov v prostredí sú skúmané v [26].

4 Kolónie gramatík

V tejto kapitole, podobne ako v predchádzajúcej, budeme venovať pozornosť systémom spolupracujúcich gramatík. Štruktúry, ktorými sa budeme zaoberať, sa nazývajú *kolónie* gramatík. Ústrednou požiadavkou pri tvorbe kolónie je požiadavka na *jednoduchosť* gramatík, tvoriacich komponenty systému a na *jednoduchosť* spolupráce komponentov. Pritom celok, *kolónia* má byť schopná vytvárať jazyky podstatne zložitejšej štruktúry ako sú jazyky jednotlivých komponentov. Základné myšlienky, vedúce ku gramatickým systémom typu kolónií, sú motivované nielen v prírode pozorovanými prejavmi činností napríklad kolónie mravcov v mravenisku, ale najmä novými požiadavkami na jednoduchosť komponentov technických systémov realizujúcich komplexné činnosti. Takouto je požiadavka na autonómnosť činnosti komponentov systémov, alebo na ich čo najjednoduchšie riadenie, požiadavka minimalizovať rozmery a váhu komponentov, taktiež možnosť jednoduchej výmeny poškodeného komponentu, alebo požiadavka na to, aby bol systém schopný ďalšej činnosti aj v prípade nefunkčnosti niektorého jeho komponentu.

Prvé práce o kolóniách gramatík boli publikované v roku 1992 [17], [18]. Množstvo informácií aj o motíváciách výskumu v tejto oblasti poskytujú práce [12-16].

Rozoberme postupne, ako v kolóniách realizujeme vyššie deklarované požiadavky na jednoduchosť komponentov a na autonómnosť ich činnosti. Venujme prvú pozornosť použitým komponentom. Kolónie používajú výlučne gramatiky schopné vytvoriť konečné jazyky. To znamená, že na úrovni komponentu systém nepracuje s iteráciou ani s rekurziou. Na vytvorenie konečných jazykov využívame (rozšírené) regulárne gramatiky, tj. gramatiky s najjednoduchším tvarom pravidiel vzhľadom na klasickú Chomského hierarchiu gramatík a jazykov.

Gramatiky kolónie nevytvárajú jednotlivo vlastné reťazce ale pracujú v spoločne zdieľanom prostredí, kde spracovávajú jeden reťazec symbolov prístupný všetkým komponentom. Jednotlivé gramatiky sa podieľajú na vytváraní reťazcov v prostredí, a to buď sekvenčne alebo paralelne podľa vopred stanoveného protokolu činnosti, ktorý je určený krokom odvodenia. Činnosť kolónie ako celku podstatne presahuje

huje možnosti poskytované jednotlivými komponentmi. Tento efekt nastáva ako dôsledok skutočnosti, že reťazce vytvárané rôznymi komponentmi môžu byť zostavené z odlišných symbolov a štartovacím symbolom gramatiky môže byť symbol produkovaný inou gramatikou. Iteratívne postupy takto nachádzajú priestor na úrovni súčinnosti komponentov.

Kolónie sú podstatne jednoduchšie ako ekogramatické systémy, aj keď analógia medzi pôsobením gramatík a akciami organizmov v spoločnom prostredí sú zrejmé.

Základný okruh problémov, ktorý táto teória rieši je stanovenie výpočtovej sily kolónií pri rozličných formách spolupráce komponentov. Správanie sekvenčných kolónií je do detailov preskúmané. Paralelné modely poskytujú zložitejšie možnosti správania aj s ohľadom na odlišné možné prístupy k riešeniu kompetenčných sporov jednotlivých komponentov. Problém nastáva, keď niekoľko komponentov má rovnaký štartovací symbol a má teda rovnaký potenciál prepísať tento symbol v prostredí a keď navyše prostredie neobsahuje dostatočný počet kópií toho symbolu. V tejto situácii sa môže javiť rovnako prirodzeným odvodenie zablokovať (pri *silnej kompetencii* komponentov), alebo pokračovať v činnosti s maximálnym počtom komponentov, ktoré sú schopné pracovať (pri *slabej kompetencii* komponentov). Stretávame sa tu s veľmi účinnými a doposiaľ v teórii s málo preskúmanými mechanizmami riadenia odvodenia alebo jeho zablokovania. Poznáme rad typických netriviálnych príkladov, ako silnú alebo slabú kompetenciu agentov využiť pre generovanie jazykov. Základný problém, porovnávajúci generatívnu silu obidvoch prístupov ostáva stále otvorený [9]. Prehľad dosiahnutých výsledkov o kolóniách je venovaná práca [25].

5 Membránové systémy

Na prelome 20. a 21. storočia boli vytvorené membránové systémy, ktoré predstavujú ďalší biologicky motivovaný model, v tomto prípade inšpirovaný biochemickými procesmi v bunkách. Model, ktorý pracuje na báze prepisovacích pravidiel rovnako, ako všetky doteraz zmienené modely, sa od menovaných gramatických štruktúr líši tým, že namiesto reťazca symbolov pracuje na membránových stromových štruktúrach. Pravidlami prislúchajúcimi k jednotlivým membránam, vrcholom membránových štruktúr, mení obsah membrán. Pravidlá, používané v membránových štruktúrach určujú zámenu jedného symbolu v membráne skupinou symbolov a navyše určujú, kde sa má vytvorený symbol, či skupina symbolov umiestniť. Do úvahy prirodzene prichádza pôvodná membrána, alebo membrána hierarchicky nad alebo pod pôvodnou membránou. Výpočet v takýchto sys-

témoch realizujeme transformáciou obsahov jednotlivých membrán. Štruktúra membrán sa môže v priebehu výpočtu meniť, napríklad sa môžu niektoré membrány rozpustiť a ich obsah sa stane obsahom hierarchicky vyššie umiestnenej membrány. O spôsoboch, akými sa môžu takéto výpočty realizovať podrobne informujú napríklad knižné publikácie [30], [4]. Niektoré typické prípady sú uvedené tiež v práci [3]. Na jednoduchom príklade načrtne činnosť membránových systémov.

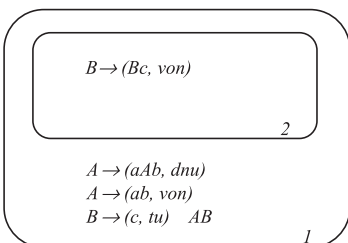
Membránový systém na obrázku 4 je systém, ktorý pracuje sekvencne, takže v každej jeho membráne je aktívne najviac jedno pravidlo. Systém obsahuje 2 membrány.

V prvej z nich sú na začiatku výpočtu objekty A a B a membrána ich dokáže upravovať troma pravidlami:

$$A \rightarrow (aAb, dnu), \quad A \rightarrow (ab, von), \quad B \rightarrow (c, tu)$$

Prvé pravidlo nahradí symbol A v prvej membráne symbolmi aAb , a celý aktuálny reťazec umiestni do membrány 2. Druhé pravidlo nahradí symbol A symbolmi ab a vysunie celý aktuálny reťazec do okolitého prostredia. Tretie pravidlo zmení B v tejto membráne na c .

Druhá membrána na začiatku výpočtu neobsahuje žiadny objekt. Pravidlom $B \rightarrow (Bc, von)$ nahradí symbol B reťazcom cB a celý aktuálny reťazec umiestni v prvej membráne. Výpočet končí, keď systém nemá možnosť prepisovať, s reťazcom obsahujúcim rovnaký počet písmen a, b, c umiestnených vo vonkajšom prostredí. Dĺžka vytvoreného reťazca



Obrázok 4

závisí od okamihu kedy prvá membrána použije pravidlo $B \rightarrow (c, tu)$.

Výpočet prebieha napríklad takto:

$$\begin{array}{ll}
 AB \text{ (1. membrána),} & aAbB \text{ (2. membrána),} \\
 aAbBc \text{ (1. membrána),} & aaAbbBc \text{ (2. membrána),} \\
 aaAbbBcc \text{ (1. membrána),} & aaAbbcc \text{ (1. membrána),} \\
 aaabbbccc \text{ (vonkajšie prostredie).} &
 \end{array}$$

Rovnako ako pri kolóniách sme mali snahu dosiahnuť čo najjednoduchšie komponenty a čo najjednoduchší spôsob komunikácie medzi nimi aj u membránových systémoch pozorujeme analogické úvahy, ktoré viedli k zavedeniu pojmu P kolónii v prácach [8], [20]. V tomto prípade výpočet prevádzajú jednomembránové komponenty

voľne umiestnené v prostredí. Pravidlá lokalizované v membránach sú štruktúrované do elementárnych programov.

Pod'akovanie: Tento príspevok vznikol v rámci riešenia grantovej úlohy VEGA 1/0692/08

Literatúra

- [1] Bartík, T.: *Monokultury a ekogramatické systémy*. Diplomová práca, Ústav informatiky, Slezská univerzita, Opava, 2007
- [2] Cienciala, L., Ciencialová, L., Kelemenová, A.: On the Number of Agents in P colonies. *Membrane Computing*. Proc. International Workshop, WCM 2007, Thessaloniki, Greece, June 2007, (G. Eleftherakis et al., eds.) LNCS 4860, Springer, Berlin, 2007, pp. 193-208
- [3] Cienciala, L., Kelemenová, A.: Život ve světě symbolů: Počítání pomocí membrán. J. Pospíchal, J. Kelemen, V. Kvasnička, eds.: *Mysel, inteligencia a život*. Vydavateľstvo STU, 2007, 369-388
- [4] Ciobanu, G., Perez-Jimenez, M.J., Paun, G.: *Applications of Membrane Computing*, Springer-Verlag, Berlin, 2006
- [5] Csuhaj-Varjú, E., Dassow, J., Kelemen, J., Păun, G.: *Grammar Systems. A Grammatical Approach to Distribution and Cooperation*. Gordon & Beach, Yverdon, 1994.
- [6] Csuhaj-Varjú, E., Kelemen, J., Kelemenová, A., Păun, G.: Eco(grammar) systems – a preview. In: *Cybernetics and Systems 94, vol. I* (R. Trappl, Ed.) World Scientific, Singapore 1994, pp. 941-948
- [7] Csuhaj-Varjú, E., Kelemen, J., Kelemenová, A., Păun, G.: Eco-grammar systems. A grammatical framework for study lifelike interactions. *Artificial Life* 3 (1997) 1-28
- [8] Csuhaj-Varjú, E., Kelemen, J., Kelemenová, A., Paun, G., Vaszil, G.: Computing with cells in environment: P colonies. *Journal of Multi-Valued Logic*, 12 3-4 (2006) 201-215
- [9] Dassow, J., Kelemen, J., Paun, G.: On Parallelism in Colonies. *Cybernetics and Systems* 24 (1993) 37- 49
- [10] Farmer, J. D., Bellin, A. d'A.: Artificial life – the coming evolution. In: *Artificial Life II, Proc. Workshop on Artificial Life* (C. G. Langton et al., eds.). Addison-Wesley, Redwood City, Cal. 1991, pp. 815-840
- [11] Kaandorp, J.A., Kuebler, J.E. : *The algorithmic Beauty of Seaweeds, Sponges and Corals*. Springer Verlag, Heidelberg, 2001
- [12] Kelemen, J.: A note on achieving low-level rationality from pure reactivity. *Journal of Experimental & Theoretical Artificial Intelligence* 8 (1996) 121-127
- [13] Kelemen, J.: Colonies as models of reactive systems. In: *New Trends in Formal Languages* (G. Paun and A. Salomaa, eds.) Springer-Verlag, Berlin, 1997, pp. 220-235
- [14] Kelemen, J.: Colonies: grammars of reactive systems. In: *Artificial Intelligence and Information-Control Systems of Robots '97* (I. Plander, Ed.) World Scientific, Singapore, 1997, pp. 27-40
- [15] Kelemen, J.: Colonies - a theory of reactive agents. In: *Proc. MFCS'98 Satellite Workshop on Grammar Systems* (A. Kelemenová, ed.) Silesian University, Opava, 1998, pp. 7-38
- [16] Kelemen, J.: A note on colonies as post-modular systems with emergent behavior. In: *Proc. Intern. Workshop Grammar Systems 2000* (R. Freund and A. Kelemenová, eds.) Silesian University, Opava, 2000, pp. 203-213

- [17] Kelemen, J., Kelemenová, A.: A subsumption architecture for generative symbol systems. In: *Cybernetics and System Research '92*, ed. R. Trappl, World Scientific, Singapore, 1992, 1529-1536
- [18] Kelemen, J., Kelemenová, A.: A grammar-theoretic treatment of multiagent systems. *Cybernetics and Systems* 23 (1992) 621– 633
- [19] Kelemen, J., Kelemenová, A.: Život ve světě symbolů: úvod do jazykově-teoretického studia životi. In: *Kognice a umělý život*. Sestavili J. Kelemen, V. Kvasnička, J. Pospíchal, Slezská univerzita v Opavě, 2001, 101-118
- [20] Kelemen, J., Kelemenová, A., Paun, Gh.: The power of cooperation in a biochemically inspired computing model: P colonies. In: *Workshop and Tutorial Proceedings, 9th International Conference on the Simulation and Synthesis of Living Systems, ALIFE IX* (M. Bedau at al., eds.) Boston, Mass., 2004, pp. 82-86
- [21] Kelemenová: Lindenmayerove systémy a ich tvorca v spomienkach, faktoch a citátoch. J. Pospíchal, P. Trebatický, J. Kelemen, V. Kvasnička, eds: *Myseľ, inteligencia a život*. Vydavateľstvo STU, 2007, 399-414
- [22] Kelemenová, A., Kelemen, J.: O biológii, matematike a výpočtoch – rozhovor s A. Lindenmayerom. *Pokroky matematiky, fyziky a astronómie* 31 (1986) 96-105
- [23] Kelemenová, A., Tupý, M.: Monocultures and homogeneous environment in eco-grammar systems. *Fundamenta Informaticae* 76 (2007) 349-365
- [24] Kelemenová, A., Vavrečková, Š.: Properties of ecocolonies. *International Journal of Foundations of Computer Science*. 2007, 17 pp. (accepted)
- [25] Kelemenová, A.: Colonies - the Simplest grammar systems. In: *Information systems and formal models*. A. Kelemenová, D. Koláří, A. Meduna, J. Zendulka (eds.), Slezská univerzita v Opavě, 2007, 157-168
- [26] Langer, M.: Generativní síla pozičních ekogramatických systémů. In: *Kognice a umělý život VI*. Sestavili J. Kelemen, V. Kvasnička, Slezská univerzita v Opavě, 2006, 259-264
- [27] Answers by A. Lindenmayer to questions of J. Kelemen and A. Kelemenová. *Bulletin of the EATCS* 23, 185-198, 1984
- [28] Lindenmayer A.: Mathematical models for cellular interaction in development I,II. *Journal of Theoretical biology* 18, (1968) 280-315.
- [29] Meinhardt, H.: *The Algorithmic Beauty of Seashells*. Springer-Verlag, New York, 2002
- [30] Păun, G.: *Membrane computing. An introduction*. Springer-Verlag, Berlin, Heidelberg., 2002
- [31] Prusinkiewicz, P., Lindenmayer, A.: *The Algorithmic Beauty of Plants*. Springer-Verlag, Berlin. 1990 (druhé vydanie 1996)
- [32] Rozenberg, G., Salomaa, A. eds.: *Handbook of formal languages I - III*, Springer Verlag, Berlin, 1997
- [33] Vavrečková, Š.: Properties of eco-colonies. In: *Information systems and formal models*, (A. Kelemenová, D. Kolář, A. Meduna, J. Zendulka, eds.) Silesian University in Opava, 2007, 235-243

Bioenergetika: kľúč k pochopeniu mozgu a mysle

Ladislav Kováč¹

Abstrakt. Prírodný život má chemickú povahu. Chémia, a nie abstraktná logika, podmieňuje, ale aj obmedzuje, jeho možnosti. Obmedzenia na evolučné možnosti života vyjadrujú dve metafory: spandrelov svätého Marka a pilot Rialta. Človek má dva rovnocenné poznávacie systémy: imunitný a nervový. Obidva fungujú na rovnakom princípe: vnútorne, autonómne, generujú variácie a prostredie z nich vhodné vyberá a stabilizuje. „Pilóty Rialta“, na ktorých je ľudský mozog evolučne vybudovaný, určujú jeho podobu na chemickej úrovni. Z komplikovanosti mozgovej chémie („biochemičtiny“) emergujú vyššie úrovne, fyziologická („fyziologičtina“) a mentálna („mentálčina“). Podľa princípu radikálnej materiálnosti kauzálne procesy mozgu bežia len na základnej, teda chemickej, úrovni a ostatné úrovne sú len jej izomorfnými, tautologickými prekladmi. Termodynamická nutnosť udržovania korelácií v komplikovanom systéme, ako aj neustále generovanie variácií robia z mozgu energeticky náročný systém, ktorý funguje naplno bez zastavenia a v podstate endogénne; na mentálnej úrovni má táto endogénna činnosť hlavnú podobu denného a nočného sna. Naše obmedzené znalosti o fungovaní nervového systému opodstatňujú tvrdenie, že 21. storočie bude storočím výskumu mozgu.

Dvadsať storočie bolo pre biológiu, podľa vyjadrenia Evelyn Fox Keller, storočím génu [43]. Šéfredaktor časopisu Science Donald Kennedy sa domnieva, že dvadsať prvé storočie by sa mohlo stať storočím mozgu [44]. Storočie génu vytvorilo predpoklady pre storočie mozgu. Predpoklady konceptuálne: to, čo vieme o génoch, možno dnes aplikovať na mozog. Je potešením konštatovať, že k pionierom takejto aplikácie patrí aj bádatelka, ktorá vzišla z našej domácej vedeckej komunity [9]. Ale aj predpoklady metodologické: zo skúsenosti storočia génu možno vyvodiť poučenie, ktoré reviduje Popperov názor na to, čo možno považovať za vedecký problém: problém sa stáva vedeckým vtedy, keď pre jeho riešenie existujú vhodné vedecké inštrumenty. Ony vymedzujú horizont vedeckého bádania. Nový prístup k mozgu otvárajú predovšetkým neurozobrazujúce techniky, ako neinvazívna komputerovaná tomografia (3D-CT-scans), pozitronová emisná tomografia (PET) a zobrazenie funkčnou magnetickou rezonanciou (fMRI), ako bude opakovane zrejme aj z údajov, na ktoré bude táto štúdia odkazovať.

Záhada mozgu, mysle a vedomia priťahovala pozornosť premýšľajúcich ľudí od nepamäti. Podľa názoru Edwarda O. Wilsona ([97], str.109) „väčšia časť dejín modernej filozofie, od Descarta a Kanta ďalej, pozostáva z neúspešných modelov mozgu“. Najlepšie modely sa vždy inšpirovali stavom vedy svojej doby a prirovnávali mozog

¹ Prírodovedecká fakulta Univerzity Komenského, Bratislava, e-mail: kovacl@fns.uniba.sk

k najkomplikovanejším strojom, aké vtedy existovali. V našej dobe modelom mozgu bol najprv analógový, potom digitálny počítač, najnovšie je to už skôr internetová sieť.

Ak nijaký z doterajších modelov neuspel, znamená to, že veda ešte na pochopenie mozgu nedorástla. Namiesto navrhovania nových modelov treba možno prístup k skúmaniu mozgu začať skromnejšie: pokúsiť sa poukázať na to, čo je v konštrukcii a fungovaní mozgu a mysle v princípe nemožné. Tým sa vymedzí priestor možností, ktorý musí akýkoľvek realistický model rešpektovať: veď tvrdenia o nemožnosti sú považované za základný fundament vedy ([24], str. 617). V tejto štúdií sa analyzujú obmedzenia, aké na mozog ukladá energetika. Už sa konštatovalo, že kognitívnu biológiu možno považovať za špecifické pokračovanie a rozvíjanie bioenergetiky [50].

Prólóg: Podstata prirodzeného života je chemická

Prirodzený život (n-life), aký vznikol a vyvinul sa na našej Zemi, má chemickú povahu. Chemická energia, teda elektromagnetická energia zviazaná s preskupovaním elektrónov na vonkajších orbitáloch atómov, je hlavnou energiou života. Predstaviteľný je však aj život, ktorý by fungoval na iných než chemických princípoch. Na obrazovkách počítačov sa darí virtuálnemu životu (v-life): prvý virtuálny život „stvoril“ už v r. 1970 John Conway a postavil ho na princípe počítačových celulárnych automatov [23]. Umelý život (a-life) vo forme sebareplikujúcich sa robotov bude čoskoro dopĺňovať, a podľa niektorých vizionárov raz aj celkom nahradí, prirodzený život. Niekde inde vo vesmíre by mohol existovať život nie v chemickej, ale elektrickej, čisto mechanickej, alebo aj v inej podobe – takej, akú si vôbec nedokážeme predstaviť.

Je podivné, ako doteraz nielen laici, ale ani vedci, a to ani biológovia a chemici, nie sú si vedomí zásadného rozdielu medzi chémiou a biochémiou. Chemické procesy v neživom svete majú skalárny charakter, bežia v čase, ale nie v priestore, zatiaľ čo chemické procesy v živých systémoch sú vektoriálne. Peter Mitchell, ktorý inherentnú vektoriálnosť biochemických procesov objavil, je pre pochopenie fungovania života nemenej významnou postavou ako je Charles Darwin [48]. Ak je Darwin Kopernikom biológie, je Mitchell Kopernikom biochémie. Nielen molekuly, náhodne, brownovsky rozložené v roztoke, ale aj ich usporiadania v priestore môžeme považovať za samostatné chemické jednotky; takými sú biologické membrány, v ktorých sú slabými, nekovalentnými, väzbami pospájané dohromady molekuly fosfolipidov, alebo proteínové siete, do ktorých sa prostredníctvom slabých väzieb organizujú molekuly proteínov. Chemickou

zmenou je nielen preskupenie kovalentných väzieb medzi atómami, ale aj adsorpcia látky na povrch membrány alebo biologickej makromolekuly, tak isto ako aj zmena polohy molekuly voči iným molekulám, napr. prenos molekuly z jednej strany membrány na druhú [73]. Neporozumieme mozgu, ak nebudeme mať túto vektoriálnu povahu biochemických procesov na pamäti.

Termodynamika určuje, čo je v prírode možné a čo nemožné. O tom, či nejaká možná štruktúra naozaj vznikne, ako rýchlo vznikne a ako dlho sa udrží, rozhoduje kinetika procesov. Vznik najpravejšej väčšiny z tých miliónov chemických zlúčenín, čo sa dnes vyskytujú na Zemi, bol favorizovaný kineticky; a nie termodynamika, ale kinetika bráni ich rýchlemu rozpadu. Pretrvávajú v nerovnovážnom, „metastabilnom“ stave: pod „tlakom“ druhej vety termodynamiky by sa korelácie medzi nimi mali postupne strácať a samé by sa mali rozpadnúť. Prírodný život ako celok predstavuje takýto „metastabilný“ stav. Mohol sa vytvoriť, rozvinúť a naďalej sa udržuje len vďaka tomu, že na Slnku, hviezde okolo ktorej naša planéta obieha, dochádza v procese jadrovej fúzie k neustálej masívnej disipácii energie. Ak by sa ľuďstvu podarilo uskutočniť riadenú jadrovú fúziu na Zemi, len by sa zmenila povaha disipatívneho zdroja, potrebného k udržovaniu a rozvíjaniu prírodného života. Termodynamika bude však ďalej vymedzovať priestor možností. Naďalej bude platiť [49], že nie všetko, čo je v prírodnom živote možné logicky, je aj prípustné termodynamicky.

Mozog má takisto chemickú povahu a jeho štruktúry, a procesy ktoré v ňom bežia, sú závislé na energii. Udržovaný a poháňaný je energiou Slnka. Pravda, sprostredkovane, cez zelené rastliny, cez metabolické procesy v zažívacom trakte, pečeni a iných orgánoch, a cez glukózu, ktorá je pre mozog hlavným zdrojom využiteľnej, chemickej, energie. Metaforicky môžeme povedať, že biochemické procesy v mozgu hovoria k nám, ľudským pozorovateľom, špecifickým jazykom, „biochemičtinou“. Tieto procesy sú organizované v čase a priestore a na vyššej úrovni analýzy ich vyjadruje iný jazyk, „fyziologičtina“. Na ešte vyššej úrovni dostáva zložitosť biochemických procesov podobu subjektívneho prežívania a vedomia, ktorému zas odpovedá špecifický jazyk, „mentálčina“. Steven Rose zdôraznil [78], že medzi biochemickým a fyziologickým popisom mozgu – a dodajme dôsledne, že aj medzi biochemickým a psychologickým popisom – nie je vzťah kauzálny, ale mapovací; ide o preklad medzi rôznymi úrovňami, teda vlastne o preklad z jedného do druhého jazyka. Konštatovanie, že chemická úroveň, interakcie molekúl, dáva jednoznačne podobu všetkým iným úrovňam na ňou, možno nazvať princípom ra-

dikálnej materiálnosti. Vyššie úrovne nemôžu kauzálne pôsobiť na nižšie; tisícročnej ilúzii, že myslenie môže ovplyvňovať či dokonca riadiť mozog, odzvonilo.

Mnohé štandardné modely mozgu fakt dominovania chémie neberú do úvahy. Jeden zo zakladajúcich otcov kognitívnych vied, George Miller vo svojom autoritatívnom prehľade o dejinách kognitívnych vied z r. 2003 medzi vednými disciplínami, ktoré podľa neho konštituuju kognitívne vedy, chémiu vôbec neuvádza [64]. Počítačové modely mozgu predpokladajú, že činnosť, akú robí mozog, možno implementovať do ľubovoľného hardwaru, takže je nepodstatné, či jej materiálnym ukotvením je biochemická štruktúra mozgu, alebo silikónový mikročip. Z profesionálnej kariéry troch pionierov modelovania mozgu sa dá priam vystopovať, prečo sa vývoj po druhej svetovej vojne uberal týmto jednostranným, a pravdepodobne pomýleným, smerom. Norbert Wiener, zakladateľ kybernetiky, sa za 2. svetovej vojny zaoberal problémom, ako zostreliť lietadlo, ktoré je v pohybe: operátora protilietadlovej zbrane považoval za súčasť riadiaceho mechanizmu a do riešenia zaviedol ideu informácie, výpočtu, spätných väzieb, vstupu a výstupu a predikcie. Tieto predstavy o riadení potom použil na vysvetľovanie činnosti mozgu. Počas druhej svetovej vojny spoznal potrebu masívnych výpočtov aj iný matematik, John von Neumann. Ukázal, že do vtedajších počítačich strojov, ktoré boli skonštruované „na pevno“, možno vložiť program a tým ich urobiť podstatne flexibilnejšími. Nečudo že potom svoju ideu počítania a programu preniesol z počítača na mozog [93]. Marvin Minsky, pionier umelej inteligencie (AI), hoci behom svojej vedeckej kariéry v mnohom menil svoje názory a svoju poslednú knihu o živote a umelej inteligencii dokonca pomenoval „Stroj emócií“, ostal verný svojmu presvedčeniu, že počítače i mozog operujú na rovnakých princípoch: aj emócie vykladá ako špecifický prípad počítania [65].

V dobe entuziastického zrovnávania mozgu s počítačom sa zabudlo na to, čo jasne vyjadril André Lwoff [59]: Za funkciami organizmov nie sú pojmy poriadku či logaritmu pravdepodobností, ale atómy a molekuly a energia žiarenia alebo chemických väzieb. Až koncom 20. storočia sa začalo zdôrazňovať, že poznávanie nielen nemožno oddeliť od jeho materiálneho substrátu, ale že v ľudskom poznávaní hrá telo podstatnú rolu a začali sa rozvíjať predstavy o „vtelenej“, inkarnovanej (*embodied*) kognícii (napr. [53], [21], [4], u nás [95]).

Je ale aj iný aspekt prirodzeného života, ktorému ani samotná bioenergetika nevenovala dostatočnú pozornosť. Asi preto, že sa sústredovala najmä len na synchronnú analýzu, nie diachrónnu: študovala mechanizmy na rôznych úrovniach organizácie súčasného života,

nevenovala sa otázkam, akým spôsobom sa tieto mechanizmy v evolúcii objavili a aký je ich význam z hľadiska fitness jednotlivých darwinovských indivíduí. O darwinovskej fitness asi len zriedkavo rozhodovala v evolúcii účinnosť premeny energie, oveľa častejšie záležalo na výkone. Z termodynamiky plynie, že maximálna účinnosť sa dosahuje vtedy, ak procesy bežia nekonečne pomaly. Organizmy, a ich stavebné moduly – teda aj mozog – boli selektované na rýchle reagovanie, na podávanie výkonu. Zvažovanie strát a ziskov, kompromis (*trade-off*), ktorým sa straty minimalizujú a zisky maximalizujú, to je to, čo charakterizuje prirodzený život.

Pilóty Rialta: evolučné obmedzenia mozgu

V „storočí génu“ sa v biológii stala postupne dominujúcou predstava organizmov ako zariadení, ktoré si „vymysleli“ molekuly nukleových kyselín, aby si „zaistili“ svoje pretrvávajúce a rozmnožovanie. Protagonistom takejto predstavy bol Richard Dawkins so svojou koncepciou „sebeckého génu“ ako replikátora a organizmov ako „vozidiel“ (*vehicles*), prostredníctvom ktorých boli sebecké gény, prípadne oportúnne združené do „gangov“, prenášané z generácie na generáciu. Bola to solídna, heuristicky užitočná predstava, ktorá vo veľkej miere určovala a podmieňovala rozvoj molekulárnej biológie a biotechnológie 20. storočia. (Aj autor tohto príspevku bol zástancom a propagátorom tejto predstavy.)

O tom, či biologická jednotka v darwinovskej súťaži s inými jednotkami pretrvá, rozhoduje jej darwinovská fitness, počet „potomkov“, ktoré vytvorí a cez ktoré ďalej pokračuje v existencii. Koncepcia „sebeckého génu“ bola podložená matematikou R.A. Fishera a J.B.S. Haldanea, ktorá elegantne počítala darwinovskú fitness nie organizmov ako evolučných jednotiek (ako ich videl Darwin), ale aliel jednotlivých génov ako samostatných, izolovaných entít. Biologická evolúcia spočívala na výbere aliel a znamenala posun vo výskyte rôznych aliel jedného génu. Takto sa dala evolúcia aj pôsobivo modelovať na počítačoch. (Prehľadom týchto metód a koncepcií je kniha J. Maynarda Smitha [61]). Ale už r. 1930 Fisherov a Haldaneov súčasník S. Wright kritizoval túto matematickú abstrakciu ako neadekvátnu, pretože „gény ktoré môžu byť výhodné v jednej kombinácii... s extrémnou pravdepodobnosťou sú nevýhodné v inej... Výber sa týka organizmu ako celku vo vzťahu k jeho prostrediu a nie jednotlivých génov ako takých“ [100]. 21. storočie, zásluhou nových inštrumentálnych techník molekulárnej biológie, ale aj nových matematických prístupov, dáva Wrightovi zapravdu. Gény sa viac nejavajú ako „fazulky vo vreci“; pospájané sú do komplikovaných sietí a už nie gén,

ba ani nie jednotlivý organizmus, ale skupina organizmov, ba aj biologický druh ako celok, sú uvažované ako jednotky v koncepcii mnohourovňovej selekcie. Génocentrizmus, zdá sa, skončí v 21. storočí v depozitári vedeckých teórií, ktoré boli vo svojej dobe užitočné, ale neskôr prekonané, rovnako ako kedysi skončil Ptolemaiov geocentrizmus.

Kritik génocentrizmu bolo po celé 20. storočie neúrekom, ale vari všetky boli bezcenné. Vo vede je zbytočné deklarovať, že svet je zložitý, že v ňom všetko súvisí so všetkým a kritizovať konkrétne modely ako zjednodušovanie, ak kritika nedokáže prísť s návrhom nových realizovateľných prístupov skúmania. Jednako, niektoré kritiky z tých čias si dnes, keď máme nové dáta a nové prístupy, zasluhujú nové čítanie. K takým patrí štúdia Goulda a Lewontina z r. 1979 [35]. Autori v článku argumentujú, že na organizmus sa treba pozeráť ako na komplikovanú stavbu (a nie, mohli by sme ich obraz rozvinúť, ako na súbor jednotlivých tehál, kde by každej tehle odpovedal jeden gén), ktorá ako celok je výsledkom darwinovského evolučného procesu, ale v ktorej početné komponenty nie sú výsledkom prirodzeného výberu, ale jednoducho nutnými dôsledkami stavebnej konštrukcie. Použili podobenstvo jedného stavebného prvku katedrály svätého Marka v Benátkach. Nad stĺpmi, ktoré držia klenbu, je kruhový oblúk (musí byť takým z konštrukčných dôvodov), no klenba je horizontálna. Takto sa objavuje pod klenbou trojuholníková plocha, označovaná ako spandrel. Je nefunkčná, ale nevyhnutná, lebo je vlastnosťou tejto konštrukcie. Môže však dodatočne dostať nejakú nezávislú funkciu, napr. estetickú, keď je na ňu namaľovaný obraz apoštolov. Termín spandrel sa dnes v biológii používa všeobecne na označenie takých komponentov buniek a organizmov, ktoré nie sú priamym produktom evolúcie prírodným výberom, ale objavili sa ako nevyhnutné dôsledky selekcie iných komponentov.

K spandrelom benátskej katedrály svätého Marka môžeme pridať iné podobenstvo, ktoré vyjadruje iné evolučné obmedzenia: pilóty Rialta. Rialto je most ponad hlavný kanál v Benátkach. Postavený je na drevených pilótach; ostatne tak, ako aj iné benátske stavby. Nosnosť pilót rozhoduje o maximálne možnej veľkosti kamenného mostu. Ak by sa most mal zásadne prestavať, muselo by sa pohnúť s jeho základmi, s pilótami. Naopak, od pilót stavba začala. Aj súčasný život je evolučne postavený na takýchto východzích „pilótach“. William Wimsatt [99] použil termín „generatívne vnorenie“ (*generative entrenchment*), aby vyjadril skutočnosť, že v biologickej evolúcii útvary, ktoré vznikli najskôr, sú základmi, na ktorých evolúcia buduje stále nové útvary, pričom tie prvé sú ponárané stále hlbšie a stávajú sa stále odol-

nejšími k akejkoľvek zmene. Evolúcia postupuje dopredu ako rohatka (ratchet), koleso so skosenými zúbkami a západkou, ktorá umožňuje otáčanie kolesa dopredu, ale znemožňuje točenie dozadu [46].

Metafora pilót Rialta vyjadruje dôležitosť východiskových, počiatočných podmienok pre celú následnú dynamiku javov. Veď to platí aj v neorganickom svete: už Newton vedel, že súčasný tvar dráhy a smer otáčania planét neplynie z gravitačného zákona, ale z východiskových podmienok. Kultúrna evolúcia človeka bude natrvalo limitovaná charakterom ľudského mozgu. Ten bol určený podmienkami, za akých došlo k oddeleniu evolúcie ľudského druhu od evolúcie iných primátov. Bolo to prostredie africkej savany: ono predstavovalo východiskové podmienky, ktoré určili špecifiku evolúcie ľudského mozgu.

Pravda, pred špecifikou mozgu bola už špecifika nervového systému jednoduchších, evolučne starších organizmov; a tá zas bola určená vlastnými východiskovými podmienkami. Nervový systém ako celok stojí na vlastných „pilótach Rialta“. Dvomi nezávislými cestami môžeme dnes preniknúť až k týmto „pilótam“. Skúmaním fylogénzy nervového systému – ako, kedy a prečo sa v biologickej evolúcii objavil – a skúmaním jeho ontogenézy. Ontogenézy preto, lebo keď sa jedinec vo svojom individuálnom vývine postupne formuje zo zygóty, ako by sa celá stavba začína odznovu a budovala na tých istých základoch, aké položila a na akých budovala biologická evolúcia.

Podľa zanechaných stôp predpokladáme, že prvé formy života existovali na Zemi už pred tri a pol miliardami rokov. Ešte pred jednou miliardou rokov sa život vyskytoval len v podobe jednobunkových organizmov. Jednobunkové organizmy nemajú nervový systém, no prvé pilóty, na ktorých je nervový systém vybudovaný, nachádzame už u nich. Prokaryotické baktérie i eukaryotické kvasinky majú molekulárne senzory, predstavované molekulami proteínov. Fundamentálnou vlastnosťou proteínov je schopnosť sentiencie [48]: prevádzanie rozmanitých konformácií s temer rovnakou energiou, z ktorých sa priviazaním liganda, látky, pre ktorú proteín funguje ako senzor, stabilizuje, a teda vyberá, len jedna. Na toto viazanie bol proteín vyselektovaný v evolúcii, ono dáva tejto interakcii liganda a senzora charakter elementárneho molekulárneho rozpoznania a robí z proteínu teleonomickú štruktúru: evolúcia zabudovala do štruktúry proteínu elementárne poznanie. Proteín tým zároveň prejavuje signifikáciu – aspektu prostredia, ktorý predstavuje ligand, pridáva význam. Rozpoznaním liganda prijíma senzor z prostredia signál a ten prenáša na inú časť svojej molekuly, takže rozpoznanie je nasledované akciou. Rozpoznanie a akcia predstavujú dohromady elementárnu kognitívnu udalosť, molekulárnu kogníciu.

U jednobunkových eukaryotických organizmov nachádzame pôvod inej pilóty nervového systému. Tieto organizmy sú schopné exocytózy: niektoré chemické látky uskladňujú v sekrečných váčkoch, z ktorých ich uvoľňujú do vonkajšieho, mimobunkového prostredia tak, že membrána váčku fúzuje s cytoplazmatickou membránou. Týmto mechanizmom sa napríklad z bunky kvasiniek uvoľňujú sexuálne hormóny [51]. Celkom tá istá molekulárna mašinéria ako v kvasinkách, len „zjemnená“ pridaním ďalších proteínov, hrá fundamentálnu rolu v prenose signálu medzi neurónmi nervového systému: neurotransmitery, uskladnené v synaptických váčkoch, sa uvoľňujú do synaptickej štrbiny, ale aj extrasynapticky, rovnakým mechanizmom, aký funguje už v kvasinkách [8]. Kvasinky majú už 20% proteínov, ktoré dnes fungujú v postsynaptických membránach myši alebo človeka; v kvasinkách majú, pochopiteľne, iné funkcie [70]. Kvasinkový feromón alfa-faktor, dekaeptid, je tak štruktúrne blízky hormónu v hypotalame, ktorý uvoľňuje gonadotropín, že tento živočíšny hormón dokáže sčasti nahradiť [58].

Posledný údaj je v súhlase s hypotézou, že hormóny i neurotransmitery sa vyvinuli z feromónov jednoduchých organizmov [84]. Keď sa jednobunkové organizmy začali meniť na mnohobunkové, s deľbou práce medzi jednotlivými bunkami, tieto látky voľne difundovali medzi bunkami a slúžili na vzájomnú komunikáciu. Bunky, čo priliehali tesne jedna na druhú, si postupne vyvinuli aj iný spôsob komunikácie: medzi membránami susediacich buniek sa vytvorili malé medzierky s proteínmi, ktoré bunky spájajú elektricky vodivými „mostíkmi“. Volajú sa trhlinové prepojenia (*gap junctions*) a existujú v rôznych tkanivách mnohobunkových organizmov. Stali sa však základňou, z ktorej sa vyvinuli najjednoduchšie prepojenia medzi nervovými bunkami – elektrické synapsy [41]. Dnes prevážna väčšina nervového systému má chemické synapsy. Možno sa však domnievať, že sa chemické synapsy objavili v evolúcii neskoršie ako elektrické. Látky, ktoré cez chemické synapsy kontakt medzi neurónmi spôsobujú, neurotransmitery, spočiatku voľne difundovali medzi bunkami a len neskôr boli uzavreté do synaptických váčkov a ich výmena medzi bunkami sa dostala pod dodatočnú kontrolu.

Súčasná štúdie ontogenézy nervového systému ukazujú, že chemikálie, ktoré v zrelom mozgu fungujú ako neurotransmitery, majú pri morfogenéze nervového systému iné poslanie: sú chemotaktickými signálmi, ktoré riadia rast a nasmerovanie axónov a ich pripojovanie na dendrity iných neurónov [6]. Z toho možno usudzovať, že toto aj bola prvá evolučná funkcia látok, ktoré dnes slúžia ako neuro-

transmitery: podieľali sa na raste, diferenciácii a morfogénéze mnohobunkových organizmov [80].

Mnohobunkovosť zvýšila fitness organizmov asi najmä tým, že došlo k oddeleniu funkcie udržovania organizmu a jeho rozmnožovania. K udržovaniu bola treba energia, a teda potrava, takže prvý typ orgánov, ktorý sa v evolúcii pre udržovanie objavil, boli ústa a zažívaci trakt. Orgány boli vystlané epitelialnymi bunkami, z ktorých niektoré sa postupne premenili na nervové bunky. Prvý nervový systém sa tvoril okolo počiatočného otvoru zažívacieho traktu, prvotných jednoduchých „úst“, najprv možno len pre ich rytmické sťahovanie a rozťahovanie [31], [34]. V prospech tejto hypotézy hovorí fakt, že aj teraz majú cicavce, človeka nevynímajúc, okrem mozgu uloženého v hlave aj druhý, ktorý býva označovaný ako „viscerálny mozog“ (*gut brain*) [33]. „Prvý“ mozog, usídlený v hlave, sa v evolúcii objavil neskoršie; vtedy, keď bolo treba riadiť už nie iba pohyb potravy zažívacím traktom, ale aj pohyb celého tela mnohobunkového organizmu, a aj jeho jednotlivých orgánov.

U človeka, ktorého mozog je v porovnaní s inými živočíchmi disproporcionálne väčší ako u iných organizmov, najmä jemné riadenie pohybov ruky si vyžiadalo veľký rast mozgu. Keď pred šiestimi miliónmi rokov klimatické zmeny v strednej Afrike prinútili predchodcu súčasného človeka opustiť bezpečnú niku konárov stromov a existovať v náročných podmienkach savany, evolučné pilóty, na ktorých spočívala jeho doterajšia evolúcia, mu už nedovoľovali uspôsobenie telesných funkcií do takej podoby, v akej ich mali jeho konkurenti, dravce a bylinožravce. Umelé pomôcky, nástroje, mohli sa stať jedinou náhradou za ostré zuby dravcov či rýchle nohy bylinožravcov. Ich zhotovovanie predpokladalo jemnú manipuláciu rukami a tá závisí od detailného riadenia pohybov ruky mozgom. Podľa viacerých bádateľov potreba jemného ovládania ruky poháňala rast mozgu a zvyšovanie ľudských kognitívnych schopností [67], [98]. Jemný pohyb ruky umožnil rozmanité gestá a možno z jazyka gest postupne evolvovala hovorená reč. Už pred 1,5 milión rokmi, 4 milióny po tom, čo sa ľudská línia oddelila od veľkých opíc, bol mozog relatívne trojnásobný väčší [11].

Veľký mozog si vynucoval veľkú lebku. Tá nakoniec dosiahla taký rozmer, že už len s ťažkosťami dokázala prejsť pri pôrode panvovým otvorom matky. Preto sa pred asi pred tristotisíc rokmi zväčšovanie lebky zastavilo – východiskové pilóty, na ktorých stojí ľudská kostra, viac nepustili [94], [88], [18], [79]. Evolúcia ľudského mozgu však pokračovala ďalej: už nie v tele matiek, ale v hlávke novorodencov. U všetkých cicavcov rastie mozog rýchlo v čase gestácie,

v tele matky, po narodení sa spomalí. Človek je výnimkou: ľudské mláďa ako by sa rodilo predčasne – odhaduje sa, že gestačné obdobie by malo trvať nie 9, ale 21 mesiacov – budovanie mozgových štruktúr pokračuje intenzívne aj po narodení (aj keď počet neurónov sa už pravdepodobne nezvyšuje), najmä v prvých mesiacoch a rokoch, a potom znovu v puberte, a celkom sa zastavuje až pri dosiahnutí veku dospelosti. To však dáva ľudskej psychickej ontogenéze jedinečný charakter: už nie iba skúsenosť, nadobudnutá v biologickej evolúcii druhu, ale aj skúsenosť získaná v kultúrnej evolúcii sa zabudováva do štruktúry mozgu a tak sa podieľa na vytváraní definitívneho repertoáru možností správania a myslenia ľudského jedinca [74]. Aj príslušníci iných živočíšnych druhov, nielen ľudský jedinec, sa učia behom individuálneho života, no základňa, na ktorom sa individuálne učenie buduje, je u človeka iná, zložitejšia, určená nielen biológiou, ale aj kultúrou. Z anatomickej a fyziologickej nevyhnutnosti je človek, ako jediný biologický druh na Zemi, ako biologickou, tak kultúrnou bytosťou. S tým samozrejme, že ľudská biológia predstavuje pilóty, ktoré podmieňujú, ale aj obmedzujú, kultúru.

Pre organizmus je jeho prostredie zdrojom potravy, teda energie a prípadne, u pohlavne sa množiacich organizmov, aj zdrojom sexuálnych partnerov – tieto zdroje organizmus potrebuje a vyhľadáva. No v prostredí sa vyskytujú aj zdroje ohrozenia: škodlivé fyzikálne a chemické faktory a popri nich príslušníci iných druhov, ale aj iní jedinci vlastného druhu – užitočné zdroje sú vzácne a s inými musí o ne súťažiť a zápasiť. Preto sa organizmus musí vo svojom prostredí orientovať a na to mu slúžia senzory. Všetky senzory sú chemickými zariadeniami (z veľkej väčšiny sú to proteíny, ale napr. aj selektívne priepustné membrány; aj registrovanie taktílnych, vizuálnych a akustických signálov sa deje chemicky). Ak najjednoduchšie organizmy, ako baktérie, majú iba desiatky typov sensorov, u človeka sú to stovky, či možno tisícky – a každý senzor zaznamenáva špecifickú vlastnosť prostredia. Nakoniec však údaje zo všetkých sensorov rezultujú len v dvoch rozdielnych druhoch akcie: v útoku (ktorým je ale aj napríklad uchopenie a prijímanie potravy, sexuálne a sociálne predvádzanie sa, dvorenie, klebetenie, klamanie, maznanie, úradovanie, vedecké bádanie, ekonomické a politické činnosti) alebo v úteku. Ohromný počet signálov musí teda organizmus, a u vyšších živočíchov ich mozog, pretransformovať iba do dvoch odpovedí. Toto je dôvod, prečo sa mozog prirovnával ku digitálnemu počítaču, ktorý na vstupoch dostáva početné dáta a na výstupe sa objavuje jediný výsledok. Zdá sa však, že biologické „pilóty Rialta“ si vynútili iný princíp fungovania: konkrétne údaje sensorov sú zovšeobecnené do binárnej podoby

„dobré“ alebo „zlé“ a toto zovšeobecnenie rozhoduje o binárnej akcii „útok“ alebo „útek“. K tomuto zovšeobecňovaniu využila evolúcia tých niekoľko desiatok chemických látok, ktoré boli predtým využívané najprv na vnútrobunkovú, neskôr na medzibunkovú komunikáciu a na morfogénu a z ktorých sa niektoré stali nakoniec aj neurotransmitermi. V podobe hormónov a emotónov sú tieto látky namiešavané do rozmanitých chemických „koktejlů“, ktoré globálne, na úrovni celého mozgu, alebo aj celého tela, určujú výslednú akciu, akú organizmus vykoná. Takto sa do dvojice senzícia-akcia v evolúcii včlenil, interkaloval, tretí komponent činnosti mozgu: emócie [47]. Na istej úrovni chemickej komplexnosti v mozgu sa vynára, emerguje, schopnosť subjektívne prežívať chemické nastavenie, ktoré má zovšeobecnú podobu „dobré“, ako príjemnosť, a to, čo vyjadruje „zlé“, ako nepríjemnosť alebo bolesť.

Dnes vieme, že okrem „drôtového prenosu“ (*wire transmission*), prenášania signálov neurónmi pospojovanými do siete cez elektrické alebo chemické synapsy, existuje v mozgu aj „objemový prenos“ (*volume transmission*) [2], [101], [85]. „Drôtový prenos“ je prenosom medzi dvomi synapsami, kým „objemový prenos“ je prenosom chemikálií, uvoľnených síce jedným neurónom, ale zasahujúcim viaceré neuróny, prípadne globálne celý mozog, cez medzibunkový priestor v mozgu (ktorý predstavuje 20% celkového objemu mozgu) a cez cerebrospinálny likvor. Liečivá, ktoré ovplyvňujú mozog, napr. anxiolytiká či antipsychotiká, ovplyvňujú mozog asi výlučne formou „objemového prenosu“, teda nepôsobia len na jeden či niekoľko málo neurónov. Zdá sa, že v prípade tých neurotransmiterov, ktoré sa podieľajú na hodnotení údajov senzorov a sú teda zviazané s emóciami, dopamínu a serotonínu, väčšina molekúl pôsobí „objemovým prenosom“ [16], [68]. Podľa pôsobivého prirovnania Agnatiho, „objemový prenos“ je čosi ako prenos rádiových vln pri vysielaní, ibaže namiesto elektromagnetického žiarenia funkciu vysielania plní difúzia chemických látok. Skôr než by sa podobal počítaču, mozog je viac ako sieť podobná špongii, ponorenej do koncentrovaného roztoku tonicky aktívujúcich chemikálií. Mozgové receptory sú ako „chemické mikrofóny“, zhotovené tak, aby verne zaznamenávali mozgovú „chemickú symfóniu“. Aj iní popisali „ektopické“ (extrasynaptické) uvoľňovanie neurotransmiterov [22], [15]. V trochu básnickom prirovnaní Terrence Sejnowki porovnal bežnú predstavu synaptického prenosu k streľbe z pušky na presne umiestnený cieľ, kým podľa jeho názoru synapsa funguje skôr ako guľovnica, čo rozprašuje guľky na všetky strany, aby zasiahli vhodné receptory uložené na rôznych miestach, aj mimo vlastnej synaptickej štrbiny [81]. Preto možno aj Vincentovo

charakterizovanie mozgu ako žľazy [91] je bližšie skutočnosti než predstava mozgu ako zariadenia podobného štruktúrou súčasným digitálnym počítačom. Už u živočíchov s najjednoduchším nervovým systémom, knidárií, existujú neurohormóny, ktoré globálne, parakrinným mechanizmom, ovplyvňujú činnosť svalov a neurónov [86]. V *Caenorhabditis elegans*, v modelovom živočíchovi iba s 302 neurónmi, dva neuróny dokážu nie cez synaptické spojenia, ale takýmto parakrinným mechanizmom (prostredníctvom najmenej 15 rozdielnych potenciálnych hormónov) ovplyvňovať metabolickú aktivitu celého tela do tej miery, že sa podstatne predlži životnosť individuálneho organizmu [10].

Prežívanie emócií v podobe príjemnosti alebo nepríjemnosti urobilo z emócií v evolúcii človeka nový evolučný faktor, ktorý podľa Darlingtona slúži pre „evolučné spevňovanie“ (*evolutionary reinforcement*) [25]. Podľa neho príjemnosť nemusí byť vždy adaptívna, výhodná z hľadiska darwinovskej fitness, no zosilňuje selekciu výhodného správania. Vyhýbanie sa trápeniam a hľadanie príjemnosti je súčasťou emocionálnej evolúcie, špecifickejšej pre človeka: honba za príjemnosťou poháňa evolúciu artefaktov a vytláča evolúciu prirodzeným výberom iným procesom, paravolúciou [47]. Naším absolutistickým vládcom je zmes chemikálií v našom mozgu; neuróny mozgovej kôry, pospájané do sietí, sú len jeho poslušnými sluhami.

Fundamentálna lekcia imunitného systému

Imunitný systém je vedľa nervového systému druhým poznávacím zariadením každého ľudského jedinca. Je nervovému systému rovnocenný: pre naše prežitie vo svete je nemenej dôležitý. Tento fakt nám uniká z jednoduchého dôvodu: činnosť svojho imunitného systému, na rozdiel od činnosti nervového systému, si neuvedomujeme. O nervovom systéme ľudstvo vie prinajmenšom dva a pol tisícročia, o existencii imunitného systému získalo prvé poznatky až na rozhraní 19. a 20. storočia.

Imunitný systém rozpoznáva a ničí antigény, látky, ktoré sú cudzie individuálnemu organizmu. Antigény rozpoznáva prostredníctvom protilátok, proteínov zo skupiny imunoglobulínov. Za jedinou sekundu stačí takto rozpoznať a zlikvidovať tisíce „nepriateľov“. Antigény sú produktmi cudzích organizmov, čo vnikli do tela a podľa bežnej predstavy, uvádzanej vo väčšine štandardných učebníc, na ničenie takýchto organizmov sa imunitný systém v evolúcii aj objavil. Hrá však nemenej dôležitú rolu v rozlíšení vlastných a cudzích proteínov (a teda v „sebaidentifikácii“) a asi aj v rozpoznávaní a odstra-

ňovaní takých vlastných proteínov alebo buniek, ktoré sa v dôsledku somatických mutácií „odcudzili“ a „zvrhli“, napr. nádorov. Rovnako ako sa nervový systém, o ktorom sa poväčšine domnievame, že dnes slúži najmä pre interakciu organizmu s jeho prostredím, vyvinul zo zariadení, ktorých prvým poslaním bolo rozpoznávanie a koordinácia činnosti vlastných buniek, aj imunitný systém mal asi podobný pôvod: podľa Jerneho a spol. jeho počiatočnou funkciou mohlo byť vnútorné rozpoznávanie a nie obrana pred cudzím [40].

Imunitný systém je schopný rozpoznať prakticky neobmedzený počet cudzích látok. Dokonca aj takých, ktoré sa v prírode nevy-skytujú a ktoré len zosyntetizovali chemici. Keď biochemici začali skúmať mechanizmus imunologickej interakcie, celkom pochopi-teľne si spočiatku predstavovali, že cudzí antigén, s presnou che-mickou štruktúrou, rigidná molekula, nejakým spôsobom vnúti formu imunoglobulínu, molekule v podstate ľubovoľne tvárniteľnej, a takto sa uskutoční rozpoznanie. Podľa teórie Linusa Paulinga [69] protilátka by mala byť bez tvaru globulou, akýsi molekulárnym vo-skom a antigén by mal fungovať ako pečiatka, ktorá do vosku otláči svoju podobu. Dnes vieme, že je tomu ináč. Do každého individuál-neho organizmu je geneticky zabudovaná schopnosť neustále gene-rovat' milióny typov protilátok, z ktorých každá má svoj jediný konkrétny molekulárny tvar. Tvarom sa myslí nielen rozmiestnenie v priestore atómov, z ktorých sa molekula skladá, ale aj rozloženie elektrických nábojov na molekule. Tých skutočných molekulárnych tvarov, nielen teoreticky možných ale v organizme naozaj prítom-ných, je tak ohromný počet, že prakticky vždy sa nájde aspoň jedna protilátka, ktorá je schopná chemicky (slabými väzbami, nie kova-lentne) interagovať s konkrétnym antigénom, dokonca aj takým, ktorý bol človekom vytvorený umelo. Môžeme bez preháňania po-vedat', že organizmus má v sebe, vo svojom imunitnom systéme, vo-pred zabudované najrozmanitejšie chemické podoby vonkajšieho sveta. Ako to charakterizoval Piattelli-Palmarini, „vôbec nič nie je 'nové' tomuto systému, repertoár existujúcich protilátok predstavuje 'sieť', interaktívny systém 'vnútorných obrazov' všetkých možných vonkajších foriem, repertoár, ktorý je 'úplný' a 'uzavretý'. Keby anti-gén, *per absurdum*, bol celkom nový pre systém, systém by s ním nemohol vôbec nič robiť. Skutočne nový antigén by bol doslova ne-viditeľný pre imunitný systém“ [71]. K tomu dodajme už len meta-forickú, ale zato fundamentálnu implikáciu: Chemický svet, ktorý v prostredí organizmu je skutočne prítomný, si sám „vyberá“ tú v organizme už zabudovanú podobu, ktorá je jeho obrazom. Anti-gén neindukuje tvorbu protilátky, neinštruuje aparát imunitného sys-

tému o tom, ako má protilátka vyzerat', ale vyberá, selektuje z existujúceho bohatého repertoáru všetkých možných protilátok takú, ktorá mu chemicky odpovedá. Organizmus sa proste už v minulosti pripravil na všetky možné alternatívy.

Potrebovať jeden-jediný typ protilátky a miesto toho syntetizovať stovky miliónov rozmanitých typov, ktoré sa nikdy nepoužijú, investovať energiu do čohosi, čo pripadá zbytočné, sa zdá byť nepochopiteľným mrhaním. Uvážme však inú alternatívu, takú, čo by bola chemicky predstaviteľná. Priestorová štruktúra proteínovej molekuly, i rozmiestnenie elektrických nábojov na nej, je určená poradím aminokyselín. V dôsledku sentencie proteínov, o ktorej bola zmienka vyššie, konformácia molekuly sa síce môže meniť, ale to vôbec neznamená, že by bola meniteľná ľubovoľne a podstatne. Obrazne povedané, molekula proteínu nemôže fungovať ako ľudská ruka, ktorá, vzhľadom k svojej flexibilitě, je schopná pevne uchopiť predmet skoro ľubovoľného tvaru. Ak by organizmus predpripravený repertoár protilátok nemal, musel by, v záujme svojho prežitia, až po vniknutí cudzieho škodlivého antigénu aktuálne skúšať naslepo syntézu jednej protilátky za druhou, kým by natrafil na správnu. Jedna molekula imunoglobulínu je zložená z asi 1 300 aminokyselín. To je 20^{1300} skúšaní, naprosto nemožné číslo. Na prehládanie celého priestoru možností bola by potrebná doba oveľa dlhšia ako je vek vesmíru, aj keby to bolo robené paralelne každou zo sto miliárd buniek čo syntetizujú protilátky (toľko B-lymfocytov má človek) s maximálnou rýchlosťou syntézy proteínov (15 aminokyselín za sekundu). A potrebná energia by odpovedala energetickému výkonu nie ľudského živočíšneho tela (okolo 100 W), ale atómovej elektrárne. Sú to teda energetické dôvody, prečo imunitný systém funguje jedinom možnom princípe: na darwinovskom princípe prirodzeného výberu. Vtedy je jeho energetická náročnosť relatívne malá [26].

Randomne tvorené rôzne druhy protilátok predstavujú darwinovské individuá, varianty, z ktorých vysokú darwinovskú fitness – teda šancu pretrvať a „rozmnožiť sa“ – má jediná, tá ktorá je „vybratá“ antigénom prostredia. Imunitný systém je však podstatne obmedzený „pilótami Rialta“, na ktorých bol evolučne vybudovaný: jeho jediný modus poznávania prostredia je chemický. Vo svojom prostredí rozoznáva jedine proteíny a prípadne nízkomolekulové látky, ktoré sú na proteín priviazané, haptény. Toto obmedzenie pomerne rýchlo vyčerpalo evolučné možnosti: hoci človek a myš mali spoločného predchodcu pred 75 miliónmi rokov, myši imunitný systém je v podstate rovnako dokonalý ako ľudský.

Energetika mozgu: Motor čo beží bez prestania na plné obrátky

Nervový systém ako poznávacie zariadenie sa odlišuje od imunitného systému v tom, že má viacero poznávacích módov. Z prostredia prijíma nie iba chemické signály, ale aj taktilné, vizuálne a akustické. Ale aj tie prekladá nervový systém rýchlo do chemickej podoby. Ak je darwinovský princíp variácií a výberu jediným energeticky prípustným princípom, na akom môže fungovať imunitný systém, možno to isté povedať o nervovom systéme?

Pionier výskumu imunitného systému Niels Jerne bol jedným z prvých, čo vyslovil domnienku, že by tomu tak mohlo byť [39]. Tento nápad rozvinuli Jean-Pierre Changeux [19] v predstave o selektívnej stabilizácii neurónových spojení v neurogenéze a Gerald Edelman [28] v koncepcii neuronálneho darwinizmu. Je známo, že pri ontogenéze centrálného nervového systému sa vytvára spočiatku nadbytok neurónov a synaptických spojení, ktorých počet sa pri „zrení“ mozgu postupne zredukuje (prehľad, aj s polemickou diskusiou je [74]). Podľa predstáv Changeuxa i Edelmana zachovávajú sa tie, ktoré sú stabilizované, a teda vlastne „vybrané“, signálmi prijímanými z prostredia. Je to proces do značnej miery podobný stabilizácii tých protilátok, ktoré si antigén „vyberie“, kým iné, ktoré v prostredí nenachádzajú „partnera“ pre chemickú interakciu, hynú.

Bežne sa u cicavcov neurogenéza mozgu mláďaťa v podstate ukončí ešte v tele matky, takže prostredie, ktoré vyberá usporiadanie neurónov a synaptické spoje medzi nimi, je vnútorným prostredím tela mláďaťa a jej matky. A základnú štruktúru určujú samozrejme gény. Ako už bola zmienka, u ľudského mláďaťa pokračuje formovanie mozgu dlho po narodení. Tam sa na ňom podieľajú signály, ktoré mláďa prijíma z vonkajšieho prostredia. V dvoch kritických obdobiach, v prvých rokoch po narodení a potom zas v puberte, prostredie zásadne ovplyvňuje vytváranú štruktúru mozgu procesom, ktorý Konrad Lorenz nazval vpečaťovaním, imprintingom [57]. Lorenz pôvodne objavil imprinting u vtákov a na vtákoch sa aj dosiaľ najčastejšie študuje; tam má v podstate jedinú podobu: fixácie mláďaťa na objekt, ktorý považuje za rodiča. Je zrejmé, že imprinting musí byť u človeka oveľa dôležitejší a univerzálnejší; je preto paradoxné, ako málo sa o ňom zatiaľ vie. „Vpečaťovanie“, „imprinting“ nie je príliš vhodné pomenovanie: navodzuje podobnú predstavu, akú mal Linus Pauling, keď sa domnieval, že proteín protilátky je voskom a antigén je pečatňou, ktorá sa otláči do vosku. To, čo neplatí o interakcii protilátka-antigén, neplatí zrejme ani v prípade mozgu: imprinting neindukuje mozgové štruktúry, neurčuje synaptické spojenia, ale vyberá z tých,

ktoré sú už k dispozícii a tie stabilizuje; nevybrané sa vymažú. Gény a imprinting v podstate napevno a natrvalo určia podobu ľudského mozgu na jeho základnej, chemickej, úrovni. Tým ale zároveň aj na vyšších úrovniach, pretože tie sú, ako už bola zmienka, len tautologickými, izomorfnými prekladmi chemickej úrovne. Teda napevno a natrvalo určia aj podobu subjektívneho prežívania a uvedomovania; vlastne podobu osobnosti.

Dôsledne domyslené, mali by darwinovské princípy neskorelovaných variácií a selekcie platiť nielen pre vývin mozgu, ale aj pre činnosť zrelého, dospelého mozgu. Tak ako imunitný systém sa „predvádza“ chemickému prostrediu antigénov bohatým repertoárom protilátok, centrálny nervový systém sa svojmu multimodálnemu prostrediu predvádza najrozmanitejšími chemickými stavmi. Patria do troch kategórií. Prvú kategórie predstavujú tie stavy, ktoré určili gény na základe fylogenetickú skúsenosti druhu. Tieto stavy sú všetkým ľuďom spoločné. Do druhej kategórie patria stavy, ktoré sa v mozgu jednotlivca vytvorili imprintingom v dvoch imprintingových kritických obdobiach. Sú špecifické pre každé individuum a behom individuálneho života sa môžu zmeniť len málo. Treťou kategóriou sú také stavy, ktoré mozog generuje aktuálne nanovo a „ponúka“ ich na vyskúšanie a otestovanie. No aj v tomto treťom prípade signály z prostredia neindukujú nič nové a sebe adekvátne, ale vyberajú z toho, čo je k dispozícii. Signál z prostredia nie je ničím, čo nesie „informáciu“, ktorú treba „spracovať“, je selektorom a spúšťačom, triggerom, vyberá z predpripravených procesov a jeden, ten vybraný, spúšťa.

Okamžite možno namietnuť: tak to nemôže byť! Či predsa svetlo z tváre, na ktorú sa pozeráme, sa neodráža do našej sietnice, kde bod po bode aktivuje fotosenzitívne receptory a nie je potom tento ohromný počet signálov prenášaný do vizuálnej kôry, kde je „počítačovo“ spracovaný a zrekonštruovaný do podoby pozorovaného objektu? Ibaže: dnes vieme, že človek má v mozgu špecifické oblasti, a špecifické neuróny, pre rozpoznávanie tváří [56], [14]. Tvár rozpoznáme okamžite a ľahko aj len z pár náznakov, pretože máme v mozgu uložený akýsi jeden modul priemernej tváre, ktorý údaje z prostredia len spúšťa a potom spresňuje. Podobne rozpoznáme rýchlo slovo hovorenej reči, lebo v mozgu už máme uložený jeho význam. Preto tiež dokážeme rýchlo a presne motoricky reagovať: údaje z fMRI naznačujú, že včasnšie, než rukou pohneme, máme už v mozgu model tohto pohybu a údaje z prostredia len tento údaj robia presnejším alebo upravujú tak, aby v budúcnosti umožnil nové rýchle konanie [38]. Práve toto triggerovanie toho, čo je už v mozgu hotové, predpripravené, by mohlo byť časťou riešenia záhady, prečo je mozog

rýchly, keď neuróny sú pomalé [9]: ako môže mozog skoro okamžite rozpoznať vizuálny objekt, ba aj ho zakategorizovať, s neurónmi ktoré vysielajú elektrické impulzy 10 miliónkrát pomalšie ako dnešný bežný počítač [87]?

Niektoré významné, či priam životne dôležité kategórie objektov alebo správania sú v mozgu uložené ako jeden modul. U jednoduchších živočíchov spočívajú poznávanie a správanie skoro výlučne na takýchto veľkých, hrubých moduloch. V evolúcii sa veľkosť modulov neustále znižovala a ich počet narastal, takže len ťažko postihneme, že princíp predvádzania hotových stavov a výberu funguje naďalej, a to aj u človeka. Dobrým príkladom je dvorenie, ako jedna z fáz sexuálneho správania [7], [62]. U drozofily dvorenie samčiekov samičkám pozostáva zo stereotypných motorických prejavov. Sú u všetkých samcov rovnaké a nasledujú za sebou v presnom poradí; ak sa ich postupnosť preruší, samček musí začať celé dvorenie celkom od začiatku. Každý krok tohto stereotypného procesu je určený špecifickým génom. Učenie môže ovplyvniť intenzitu dvorenia, ale nie jeho sekvenciu. Sexuálne dvorenie u človeka je oveľa menej stereotypné, ale je zrejme, že je evolučne vybudované na tých istých pilótach, ako dvorenie drozofily. Len moduly, z ktorých ľudské dvorenie pozostáva, „kvádre“, z ktorých je postavené, sú oveľa menšie a je ich podstatne väčší počet. To to dáva celému správaniu onú flexibilitu, ktorú od dávna popisovali a obdivovali umelci. Gény pre dvorenie sú už u drozofily usporiadané hierarchicky, pričom jednému z nich, „veliteľskému“ (*master gene*), sú podriadené ostatné. Ale rovnako je tomu pri ľudskom dvorení: pár génov vysoko v hierarchii, na ktorých je závislá neuroendokrinná aktivita hypotalamu, má celkom pod kontrolou kôrové procesy, od „vôľových“ konaní, ktoré si jedinec prípadne len do datočne racionalizuje, až po extaktické výlevy básnikov.

To isté platí pre vizuálne poznávanie. Podobne ako má človek v mozgu podobu „všeobecnej“ tváre, má aj iné predpripravené veľké moduly. Väčšina vizuálneho poznávania je však výberom z veľkého počtu veľmi jemných modulov. Aj u živočíchov jednoduchších ako je človek. Klasické štúdie Hubela a Wiesel na mačkách ukázali, že vizuálny systém má hotové moduly horizontálnych a vertikálnych tyčiek a hrán, farieb či iných vlastností [96]. Z nich sa potom skladá dohromady konečný obraz objektu – akoby podoby, ktorú má objekt vo vonkajšom prostredí, v skutočnosti však skonštruovanú v mozgu z „pre-fabrikovaných“ modulov. Podľa Fisera a spol. „v priebehu senzorického kódovania aktivita vo vizuálnej kôre, spustená stimulom, v podstate odráža modulovanie a trigerovanie dynamiky vnútorných obvodov senzorickými signálmi a nie že by priamo kodovala štruk-

túru vstupujúceho signálu“ [30]. Podobne z fMRI na slepých ľuďoch Burton a spol. dospeli k záveru, že mozog operuje vnútorne, sám od seba, a senzorické údaje len modulujú tieto operácie [17]. Ved' preto aj dráždenie špecifických oblastí mozgovej kôry vnímajú slepci ako svetelné záblesky a dráždenie iných oblastí vnímajú hluchí ako hluk. Mohli by sme parafrázovať Platóna, že celá naša mentálna činnosť je len „rozpomínaním sa“ na to, čo nám do mozgu vložila evolúcia; dokonca aj náš repertoár pojmov môže byť v nás preformovaný a behom života si ho už len upresňujeme [50]. Naša schopnosť bohatého a detailného pojmového chápania javov len odráža našu schopnosť uchopiť rukou predmety takmer ľubovoľných tvarov, vďaka výnimočnej flexibilitě ľudskej ruky.

Dnes hádam jedným z najpresvedčivejších dôkazov, že v prostredí poznáme len to, na čo sme vopred vnútorne nastavení, je vysvetlenie ľudskej schopnosti empatie cez činnosť zrkadliacich neurónov [1]. To, čo cíti náš blížny, poznáme len vtedy, keď sledovanie radosti alebo bolesti iného človeka aktivuje v našom mozgu tie isté oblasti, ako keby sme tie isté pocity prežívali my sami. Psychopati, ktorým sa príslušné oblasti mozgu neaktivujú a teda v sebe nemajú „predpripravené“ poznanie, nie sú schopní empatie.

Udržovanie mozgových usporiadaní, aké boli nastavené v evolúcii (a sú „predpísané“ génmi) alebo v ontogenéze mozgu (imprintingom) a generovanie nových alternatívnych stavov sa podobá udržiavaniu stavov imunitného systému (imunologická pamäť) a generovaniu nových protilátok a vyžaduje si trvalý prísun zdrojov energie. Z rovnakých dôvodov, z akých je energeticky nemožné, aby imunitný systém „ušil na telo“ jednému antigénu odpovedajúcu protilátku, pretože by bolo treba podať „astronomicky“ veľký výkon, je energeticky nemožné, aby mozog robil akýkoľvek kognitívny akt od začiatku, *ab initio*.

Pritom všetkých mozgových stavov, či už určených dedičnosťou, imprintingom alebo aktuálnym varírovaním, je zrejme oveľa viac ako je stavov imunitného systému. Preto by sa dalo čakať, že mozog bude energeticky náročnejší, ako je imunitný systém. A naozaj. Je všeobecne známo, že ľudský mozog, hoci váži len okolo 2% ľudského tela, sa na spotrebe chemickej energie telom podieľa 20% percentami (ba v prvom roku života je to viac ako 60%; u dospelých opíc len 8%). V prepočte na jeden gram spotrebováva ľudský mozog toľko energie ako srdcový sval [3]. Je to zhruba šestnásťkrát viac ako je spotreba kostrového svalu v pokoji [55], alebo toľko, koľko spotrebovávajú svaly nôh pri maratónskom behu [5]. Krv, ktorá ľudským mozgom preteká – viac ako pol litra každú minútu – prináša nielen zdroje che-

mickej energie, ale je aj chladiacou kvapalinou: vysoký výkon mozgu potrebuje účinné chladenie [29]. Prečo je potreba energie v mozgu takáto vysoká?

Dôkladnú energetickú bilanciu mozgu urobili Attwell a Laughlin [5]. Svoju analýzu obmedzili na šedú kôru mozgu, v ktorej je 90% všetkých neurónov excitačných a ich neurotransmitterom je glutamát. Podľa ich výpočtov 85% chemickej energie kôry ide na nervový prenos. Tieto údaje, platné pre neokortex krysy, boli prepočítané s podobným výsledkom pre neokortex človeka [54].

Údaje Ames, počítané na celý mozog, sú podobné [3]. Len 5-15% chemickej energie sa spotrebáva na „bežné“ chemické procesy, aká majú bunky mozgu spoločné s inými bunkami. 40-50% ide na prenos Na^+ iónov membránami, 3-7% na transport Ca^{2+} , 10-20% na syntézu a transport neurotransmitterov, ich umiestnenie do synaptických váčkov, exocytózu do medzibunkového priestoru a retransport, 20-30% na vnútrobunkovú signalizáciu (aktiváciu a dezaktiváciu proteínov napr. fosforylovaním, a tvorbu a fungovanie druhotných poslov, ako cAMP, cGMP, inozitolových zlúčenín) a 20-30% na transport pozdĺž axónov a dendritov a prestavbu cytoskeletu. V súhlase s tým sú aj merania s magnetickou rezonančnou spektroskopiou [83], podľa ktorých 80% aktivity ľudskeho mozgu je viazané s cyklovaním glutamátu ako hlavného excitačného transmittera mozgovej kôry. Pritom táto aktivita je vysoká aj v mozgu, ktorý neprijíma nijaké vonkajšie stimuly; stimulácia ju zvýši zhruba iba o 10% [82].

To, že neuróny kôry javia spontánnu aktivitu aj bez vstupov z prostredia sa vedelo dávno. Väčšinou sa ale vysvetľovalo ako prejav šumu, analogicky ako aj uzavreté membránové kanály javia molekulárny šum. Ukázalo sa však, že ide o aktivity korelované medzi populáciami neurónov v rozsiahlych oblastiach kôry [89]; interpretovalo sa to tak, že v neprítomnosti vonkajšieho stimulu neuróny kôry „blúdria“ po rozmanitých stavoch a vonkajší stimul ich posunie do jedného z takýchto stavov, adekvátneho príslušnému stimulu [60]. Na mentálnej úrovni sa to prejavuje ako „blúdenie“ mysle, ktorej zmyslom by mohlo byť udržiavanie optimálnej bdlosti, navodzovanie pocitu koherencie minulého, prítomného a budúceho prežívania alebo jednoducho delenie pozornosti medzi rozmanité mentálne úlohy [63], [72]. Čakalo by sa, že mentálnej záťaži, napr. riešeniu zložitej matematickej úlohy, by mala odpovedať rovnako veľká chemická záťaž a malo by teda dôjsť ku zvýšeniu metabolizmu i spotreby chemickej energie mozgom. Fakt, že sa pri meraní energetickej bilancie celkového mozgu nepozorovali rozdiely medzi pokojovým, „odpočívajúcim“ a medzi „zaťaženým“ mozgom bol preto dlho záhadou [76].

Záhada sa vyriešila, keď sa zobrazovacími technikami začali merať lokálne toky energie v jednotlivých častiach mozgu. V tej oblasti mozgu, ktorá sa angažuje v nejakej mentálnej činnosti, nielen v riešení matematickej úlohy, ale napr. aj pri pozeraní fotografií, hraní na hudobný nástroj a pod., sa zvyšuje prietok krvi a spotreba kyslíka a glukózy – zrejmý znak zvýšenej biochemickej aktivity a teda aj zvýšenej spotreby chemickej energie. Ak sa aktivuje jedna oblasť, relatívne menej je aktívna iná oblasť [36].

V analógii s „tmavou energiou“ vesmíru, o ktorej zatiaľ skoro nič nevieme, nazval Marcus Raichle tú energiu, ktorá v mozgu neslúži na „spracovanie“ údajov z prostredia, ale sa využíva na intenzívnu vnútromozgovú (a vnútromentálnu) činnosť „tmavou energiou mozgu“ [76], [75]. Raichle uvažoval dve možnosti, čomu slúži táto vysoká endogénna aktivita, ktorú nazval defoltovou [77]. Jedna možnosť, podložená teoretickými výpočtami, by bola, že v mozgu sa udržuje rovnováha medzi neustále fungujúcimi excitačnými a inhibičnými synapsami, a synapsy v tomto stave labilnej rovnováhy – mohli by sme módne povedať, na okraji chaosu – javia bohatú dynamiku a môžu na vonkajšie stimuly reagovať rýchlejšie ako je rýchlosť jednotlivej synapsy. (V zmysle konceptu sentiencie [48] by sme mohli dodať, že by táto bohatá dynamika predstavovala špecifickú sentienciu na hierarchickej úrovni mozgu.) Pri druhej možnosti by vnútorná činnosť mozgu mohla podľa Raichla spočívať v udržovaní informácie, potrebnej na reagovanie, alebo aj na predikovanie podnetov z prostredia. Mozog by mohol fungovať ako bayesovský inferenčný stroj. Iní uvažujú mimo iné aj možnosť, že táto neustála endogénna činnosť by mohla slúžiť konsolidácii a stabilizovaniu pamäti [63]. Terrence Sejnowski sa domnieva, že vyjasnenie vysokej spontánnej aktivity mozgu by mohlo byť cestou k poznaniu povahy vedomia a pocitu „ja“ [81]. Treba však mať na pamäti, že vysoká spontánna aktivita v tých častiach mozgu, ktoré sú u človeka spojené s porozumením mentálneho stavu iných ľudí, morálnym argumentovaním, premýšľaním o sebe samom a plánovaní budúcnosti sa pozorovala aj v mozgoch opíc v hlbokoj anestéze, teda určite v úplnom bezvedomí [92]. To podporuje názor, že o činnosti mozgu – a teda implicitne aj o mentálnych funkciách – rozhoduje na prvom mieste jeho vnútorná biochemická dynamika [72].

Je známo, že mozog je počas spánku temer rovnako aktívny, ako v bdelom stave. Veľkú časť spánku zaplňujú sny. Existujú najrozmanitejšie teórie a zmysle a funkcie nočného sna. Pováčšine sa súdi, že sa behom sna, tak ako spánku vôbec, konsoliduje pamäť na udalosti z predošlého dňa. Podľa Allana Hobsona nočný sen funguje podobne

ako imunitný systém: konštruuje najrozmanitejšie situácie, z ktorých ale len niektoré sa prejavia ako užitočné v skutočnom živote [37]. Možno teda vysloviť hypotézu, že skôr než usporiadúvanie minulosti je nočný sen prípravou na možné budúce opcie: bol by do značnej miery náhodným, ale pritom skordinovaným, nastavovaním synaptických spojení a hladín emotónov. To isté však možno povedať aj o dennom sne. Eric Klinger je presvedčený, že ľudská myseľ venuje polovicu času v bdelom stave dennému sneniu [45]. Myslí sa tým nielen fantazировanie ako pri nočnom sne – pri dennom snívaní najmä v podobe imaginárnych emocionálnych uspokojení – ale aj analýza minulých činností a plánovanie budúcich, premýšľanie o možných i nemožných alternatívach – všetko, samozrejme, vždy v podobe súvislých príbehov. Aj tu zrejme na základnej, materiálnej úrovni mozgu ide o chemické procesy podobné tým, aké trvalo bežia v našom imunitnom systéme – rozdiel je len v tom, že sme ich schopní na mentálnej úrovni zaznamenať a konceptualizovať.

Táto mentálna endogénna činnosť, preložená z „mentálčiny“ do „biochemičtiny“, má aj na biochemickej úrovni podobu konzistentných chemických „príbehov“. V hovorom jazyku by sme ju ľahko nazvali „spontánnou“, no ona termodynamicky spontánna nie je; naopak, je hlavným spotrebičom chemickej energie mozgu. Keby nebežala, v komplikovanej stavbe mozgu by sa rýchlo strácali korelácie (v dôsledku druhej vety termodynamiky!) a v systéme by termodynamicky spontánne narastala neusporiadanosť. Takto mozog pripomína auto, ktorého prevádzkyschopnosť sa udržuje tým, že má motor stále zapnutý, aj keď vozidlo stojí. Krv, ktorá mozgom intenzívne preteká, pri naša nielen pohonné látky a mazací olej, ako je tomu v dvojtaktnom motore, ale je zároveň chladiacom kvapalinou. Vysoká endogénna aktivita zároveň zaisťuje, že mozog je v trvalej pohotovosti a na vonkajšie stimuly dokáže promptne zareagovať.

Epilóg: Neadekvátnosť počítačovej metafory mozgu

Udržovanie a generovanie stavov mozgu predstavuje sled početných chemických procesov, prebiehajúcich v zložito štruktúrovanom systéme. Ak sú dva chemické procesy rovnaké, ale bežia v rozdielnych štruktúrnych kontextoch, predstavujú samostatné chemické udalosti. Takže v mozgu beží v každom okamžiku ohromné množstvo paralelných chemických udalostí. Väčšina z nich je endergonická, spotrebáva chemickú energiu. Bádatelia, čo mozog chápu ako počítač, miesto o chemických udalostiach hovoria o elementárnych počítačových operáciách. Takto spočítali, že mozog vykoná za jednu sekundu okolo 10^{15} operácií. Odhaduje sa, že už na 12 rokov by mohli byť na

trhu lacné stolné počítače, ktoré sa výpočtovou kapacitou vyrovnajú ľudskému mozgu [66].

Úvaha je to celkom pomýlená. Ak je mozog strojom, je chemickým strojom, jednocelovým. Jeho jediným poslaním je zaistiť udržanie a rozmoženie svojho nositeľa, individuálneho organizmu. Chemické procesy, ktoré v mozgu bežia, sú jeho sémantikou. Naproti tomu počítače sú univerzálnymi, syntaktickými strojmi (v duchu Turingových predstáv), určenými na prevedenie ľubovoľných výpočtov na základe programov a vstupov, ktoré do nich vkladá ľudský subjekt. Sú ľudskými exosomatickými orgánmi. Počítač spracuje vstupy a po ukončení programu sa v činnosti zastaví – aký je to diametrálny rozdiel od mozgu, ktorý stále, deň a noc, beží „na plné obrátky“! Roboty, ktoré budú vybavené tak výkonnými počítačmi, že budú schopné sebareplikácie a budú (ak sebareplikáciu považujeme za charakteristiku života) živé, nebudú mať s prirodzeným životom, ktorého súčasťou je aj ľudský druh, vôbec nič spoločné.

Tento záver možno podprieť skúsenosťou z dvoch etáp doterajšieho vývoja umelej inteligencie (AI), ktorá mala byť napodobnením činnosti mozgu [52]. Pôsobivo ich na svojom osobnom príklade popísal Rodney Brooks [13]. Prvá fáza začala pred viac ako 50 rokmi, dnes sa považuje za prekonanú a obyčajne sa označuje, podľa návrhu Johna Haugelanda z r. 1985, termínom GOFAI („good old-fashioned AI“). Vychádzala z predstavy reprezentačnej teórie mysle. Podľa nej inteligentný živý tvor je riadený myslou ako počítačom, ktorý robí výpočty nad mentálnymi reprezentáciami vonkajšieho sveta. Robot mal byť čímsi ako kópiou takéhoto živého tvora. V GOFAI robot dostáva od programátora hneď na začiatku úplný popis prostredia a explicitný zoznam inštrukcií a pravidiel, podľa ktorých počíta. Namiesto toho v NAI („new AI“), ktorej je dnes Brooks jedným z najprominentnejších predstaviteľov, robot je „situovaný“ vo svojom prostredí. V prostredí sa správa reaktívne, bez vopred pripraveného programu. Reakcie, ktoré sa osvedčili, si ukladá do pamäti, takže sa sám postupne učí; k tomu programátora nepotrebuje. Podľa Brooksa kognitívne systémy nie sú pasívnymi počítačmi, ale agentmi. Nemajú nijakú centrálnu riadiacu jednotku, ich inteligencia sa buduje postupne vo veľkom počte paralelných procesov, ktoré sú medzi sebou len voľne pospájané a koordinované interakciami s prostredím. Súčasťou prostredia sú aj iní agenti (a či skôr iné agenty?). (U nás NAI pekne popísal Kelemen [42].) Nedávno s pôsobivou kritikou do vlastných radov prispel profesor informatiky z Yale David Gelernter [32]. Mozog podľa neho nie je počítačom: počítač je ľahostajný k tomu, aké inštrukcie vykonáva, podobne ako sa rúra v peci nezaujíma o to,

čo sa v nej pečie. To nie je náš prípad: cez mozog prežívame svet, prečítujeme emócie, chémia nás robí inými, než sú počítače.

Hádam nič neilustruje lepšie nesúmerateľnosť chemických procesov, na ktorých je postavený prirodzený život, a digitálneho počítačania, ako proces zvinutia (*folding*) proteínov. Jednoduchý proteín sa spontánne zvinie do svojej natívnej štruktúry behom milisekúnd, alebo nanajvýš sekúnd. V procese sa minimalizuje voľná energia interakcií medzi aminokyselinami, ktorých sekvencia predstavuje jeho primárnu, denaturovanú štruktúru. Ale na výkonných digitálnych počítačoch sa zatiaľ darí simulovať len prvú kratulinku častí prirodzeného procesu zvinutia, niekoľko desiatok nanosekúnd, o šesť rádov menej v porovnaní s tým, čo „dokáže“ príroda. Aby sa výkonnosť simulovania zvýšila, zakladajú sa medzinárodné združenia, ktorých poslaním je zapojiť do počítania prostredníctvom internetu desaťtisíce počítačov na celom svete: kým stolný počítač nie je využívaný jeho užívateľom a „odpočíva“, môžu v ňom bežať programy čo spracovávajú dáta pre počítanie, doručené koordinátorom simulovania (napr. [27]). Nielen sa spotrebovávajú veľa času, ale aj ohromné množstvá elektrickej energie na simulovanie čohosi, čo v prírode beží rýchlo a termodynamicky spontánne. Proteín proste nie je digitálny počítač, aj keď býva často k nemu prirovnávaný, ba dokonca aj zaň považovaný.

Stotožňovať mozog s počítačom znamená očakávať od neho ten istý nemožne veľký počítačový výkon, aký podávajú všetky cez internet združené počítače, keď simulujú zvinutie i len jediného jednoduchého proteínu. Skôr než digitálnym počítačom je mozog, podobne ako proteín, dynamickým systémom [20], [90]. Ako má proces zvinutia proteínov atraktor, ku ktorému smeruje – minimum voľnej energie – aj v mozgu sú možno prítomné rozmanité atraktory, niektoré dané nomicky, prírodnými zákonmi, niektoré teleonomicky, génmi, imprintingom, možno ďalšími, doteraz neobjavenými faktormi. Spôsob, akým ľudský mozog rozpoznáva konkrétnu tvár cez zabudovaný prototyp univerzálnej tváre (tento prototyp môžeme považovať za atraktor), je možno paradigmatickým príkladom.

Z argumentov, ktoré boli v tejto štúdie uvedené, je zrejmé, že ani modely GOFAI, ani NAI nevystihujú spôsob, akým funguje ľudská myseľ, a teda aj mozog. Nemožno povedať, že GOFAI a NAI predstavujú dva protikladné, krajné, ideálne typy a ľudský mozog sa nachádza niekde medzi nimi. Jednoducho zatiaľ nijaký plauzibilný model nemáme. Rodney Brooks to asi presne vystihol, keď medzi možnosťami, prečo sme zatiaľ neúspešní, menoval aj túto: „Naším modelom biológie možno chýba niečo fundamentálne a zatiaľ nepredstaviteľné“ [12]. To nie je ale podnet k rezignácii, ale k ďalšiemu

intenzívneho bádaniu. Môžeme vymedzovať nemožné – to bolo primárnym cieľom tohto príspevku. V rámci možného treba však predkladať i skúšať čo najviac alternatív. Práve tak, ako to robia obidva prirodzené poznávacie systémy, ktorými disponujeme, imunitný a nervový systém. Čím viac alternatív bude k dispozícii, tým väčšia bude pravdepodobnosť, že sa medzi nimi vyskytne aj taká, ktorá bude odpovedať skutočnosti a bude teda rozšírením nášho poznania.

Literatúra

- [1] Abbott, A. (2007) Scanning psychopaths. *Nature* 450, 942-944.
- [2] Agnati, L.F., Bjelke, B., Fuxe, K. (1995) Volume versus wiring transmission in the brain: a new theoretical frame for neuropsychopharmacology. *Med. Res. Rev.* 15, 33-45.
- [3] Ames, A. (2000) CNS energy metabolism as related to function. *Brain Res. Rev.* 34, 42-68.
- [4] Anderson, M.L. (2003) Embodied cognition: A field guide. *Artificial Intelligence* 149, 91-130.
- [5] Attwell, D., Laughlin, S.B. (2001) An energy budget for signaling in the grey matter of the brain. *J. Cerebral Blood Flow and Metabolism* 21, 1133-1145.
- [6] Bagnard, D. (ed.) (2007) Axon guidance and growth. Springer, New York.
- [7] Baker, B.S., Taylor, B.J., Hall, J.C. (2001) Are complex behaviors specified by dedicated regulatory genes? Reasoning from *Drosophila*. *Cell* 105, 13-24.
- [8] Bennett, M. K., Scheller, R.H. (1993) The molecular machinery for secretion is conserved from yeast to neurons. *Proc. Natl. Acad. Sci. U.S.* 90, 2559-2563.
- [9] Beňušková, L., Kasabov, N. (2007) Computational neurogenetic modeling. Springer, New York.
- [10] Bishop, N.A., Guarente, L. (2007) Two neurons mediate diet-restriction-induced longevity in *C. elegans*. *Nature* 447, 545-549.
- [11] Bradbury, J. (2005) Molecular insights into human brain evolution. *PLoS Biology*, 3, 367-370.
- [12] Brooks, R.A. (2001) The relationship between matter and life. *Nature* 409, 409-411.
- [13] Brooks, R.A. (2002) *Flesh and machines: How robots will change us*. Pantheon, New York.
- [14] Bruce, V., Young A. (1998) *In the eye of the beholder: The science of face perception*. Oxford Univ. Press, Oxford.
- [15] Bunin, M.A., Wightman, R.M. (1998) Quantitative evaluation of 5-hydroxytryptamine (serotonin) neuronal release and uptake: An investigation of extrasynaptic transmission. *J. of Neurosci.* 18, 4854-4860.
- [16] Bunin, M.A., Wightman, R.M. (1999) Paracrine neurotransmission in the CNS: Involvement of 5-HT. *Trends in Neurosci.* 22, 377-382.
- [17] Burton, H., Snyder, A.Z., Raichle, M.E. (2004) Default brain functionality in blind people. *Proc. Natl. Acad. Sci. U. S.* 15500-15505.
- [18] Calvin, W.H. (1991) *The ascent of mind*. Bantam Books, New York.
- [19] Changeux, J.-P. (1983) *L'homme neuronal*. Fayard, Paris.
- [20] Carello, C., Turvey, M. T., Kugler, P. N. and Shaw, R. E. (1984) Inadequacies of the computer metaphor. V: Gazzaniga, M. S. (Ed.) *Handbook of cognitive neuroscience*. Plenum Press, New York, pp. 229-248.
- [21] Clark, A. (1997) *Being there: Putting brain, body and world together again*. MIT Press, Cambridge, MA.
- [22] Coggan, J.S., Bartol, T.M., Esquenazi, E., Stiles, J.R., Lamont, S., Martone, M.E., Berg, D.K., Ellisman, M.H., Terrence J. Sejnowski, T.J. (2005) Evidence

- for ectopic neurotransmission at a neuronal synapse. *Science* 309, 446-451.
- [23] Csontó, J., Palko, M. (2002) *Umělý život*. Elfa, Košice.
- [24] Daly, H.E., Townsend, K.N. (1993) *Valuing the earth: Economy, ecology, ethics*. MIT Press, Cambridge, MA.
- [25] Darlington, P.J. (1975) Group selection, altruism, reinforcement, and throwing in human evolution. *Proc. Natl. Acad. Sci. U.S.* 72, 3748-3752.
- [26] Derting, T.R., Compton, S. (2003) Immune response, not immune maintenance, is energetically costly in wild white-footed mice (*Peromyscus leucopus*). *Physiol. Biochem. Zool.* 76, 744-752.
- [27] Duan, Y., P. A. Kollman, P.A. (2001) Deep computing for the life sciences. *IBM Systems Journal* 40, 297-309.
- [28] Edelman, G.M. (1987) *Neural darwinism: The theory of neural group selection*. Basic Books, New York.
- [29] Falk, D. (1992) *Braindance*. Holt, New York.
- [30] Fiser, J., Chiu, C., Weliky, M. (2004) Small modulation of ongoing cortical dynamics by sensory input during natural vision. *Nature* 431, 573-578.
- [31] Galliot, B., Miller, D. (2000) Origin of anterior patterning: how old is our head? *Trends in Genet.* 16, 1-5.
- [32] Gelernter, D. (2007) Artificial intelligence is lost in the woods. *MIT Technology Review*, July 2007.
- [33] Gershon, M. (1999) *The Second Brain*. Harper Perennial, New York.
- [34] Ghysen, A. (2003) The origin and evolution of the nervous system. *Int. J. Dev. Biol.* 47, 555-562.
- [35] Gould, S.J., Lewontin, R. C. (1979) The spandrels of San Marco and the panglossian paradigm: a critique of the adaptationist programme. *Proc. Royal Soc.* 205, 851-998.
- [36] Greicius, M.D., Krasnow, B., Reiss, A.L., Menon, V. (2003) Functional connectivity in the resting brain: A network analysis of the default mode hypothesis. *Proc. Natl. Acad. Sci. U.S.* 100, 253-258.
- [37] Hobson, A. (1994) *The chemistry of conscious states*. Little and Brown, New York.
- [38] Ito, M. (2000) Internal model visualized. *Nature* 403, 153-154.
- [39] Jerne, N.K. (1968) Antibodies and learning: selection versus instruction. V: Quarton, G.C., Melnechuk, T., Schmitt, F.O. (eds.) *The neurosciences: A study program*. Rockefeller University Press, New York.
- [40] Jerne, N.K., Roland, E., Cazenave, P.-A. (1982) Recurrent idiotypes and internal images. *The EMBO J.* 1, 243-247.
- [41] Kandel, E.R., Schwartz, J.H., Jessell, T.M. (2000) *Principles of neural sciences*. McGraw-Hill, New York.
- [42] Kelemen, J. (1994) *Strojovia a agenty*. Archa, Bratislava.
- [43] Keller, E.F. (2000) *The century of the gene*. Harvard University Press, Cambridge, MA.
- [44] Kennedy, D. (2000) Triple play. *Science* 290, 709.
- [45] Klinger, E. (1990) *Daydreaming*. Tarcher, Los Angeles.
- [46] Kováč, L. (2000) Fundamental principles of cognitive biology. *Evolution and cognition* 6, 51-69.
- [47] Kováč, L. (2003) Ľudské vedomie je produktom evolučnej eskalácie emocionálneho výberu. In: (Kelemen, J., Ed.) *Kognície a umělý život*. III. Slezská univerzita, Opava, pp. 75-93.
- [48] Kováč L., (2006) Life, chemistry and cognition. *EMBO Rep.* 7. 562-566.
- [49] Kováč, L., Nosek, J., Tomáška, E. (2003) An Overlooked Riddle of Life's Origins: Energy-dependent Nucleic Acid Unzipping. *J. Mol. Evol.* 57, S182-S189
- [50] Kováč L., Rybár, J. (1994) Úsilie o exaktnú epistemológiu. *Organon F* 1, 133-141

- [51] Kurjan, J. (1992) Pheromone response in yeast. *Ann. Rev. Biochem.* 61, 1097–1129.
- [52] Kvasnička, V., Pospíchal, J. (2002) Konekcionizmus a modelovanie kognitívnych procesov, str. 257-345. V: Rybár, J., Beňušková, Ľ., Kvasnička V. (eds.) *Kognitívne vedy*. Kalligram, Bratislava.
- [53] Lakoff, G., Johnson, M. (1999) *Philosophy in the flesh: the embodied mind and its challenge to Western thought*. Basic Books, New York.
- [54] Lennie, P. (2003) The cost of cortical computation. *Curr. Biol.* 13, 493-497.
- [55] Leonard, W.R., Robertson, M.L., Snodgrass, J.J., Kuzawa, C.W. (2003) Metabolic correlates of hominid brain evolution. *Compar. Biochem. Physiol. A* 136, 5-15.
- [56] Leopold, D.A., Bondar, I.V., Giese, M.A. (2006) Norm-based face encoding by single neurons in the monkey temporal cortex. *Nature* 442, 572-575.
- [57] Lorenz, K. (1966) *On Aggression*. Methuen, London.
- [58] Loumaye E, Thormer J, Catt K. J. (1982) Yeast mating pheromone activates mammalian gonadotrophs: Evolutionary conservation of a reproductive hormone? *Science* 218, 1323-1325.
- [59] Lwoff, A. (1962) *Biological order*. MIT Press, Cambridge, MA.
- [60] Mason, M.F., Norton, M.I., Van Horn, J.D., Wegner, D.M., Grafton, S.T., Macrae, C.N. (2007) Wandering minds: The default network and stimulus-independent thought. *Science* 315, 393-395.
- [61] Maynard Smith, J (1989) *Evolutionary genetics*. Oxford University Press, Oxford.
- [62] Mehren, J.E., Ejima, A., Griffith, L.C. (2004) Unconventional sex: fresh approaches to courtship learning. *Curr. Opin. Neurobiol.* 14, 745-750.
- [63] Miall, R.C., Robertson, E.M. (2006) Functional imaging: In the resting brain resting? *Curr. Biol.* 16, R998-R1000.
- [64] Miller, G.A. (2003) The cognitive revolution: a historical perspective. *Trends in Cogn. Sci.* 7, 141-144.
- [65] Minsky, M. (2006) *The emotion machine: Commonsense thinking, artificial intelligence, and the future of the human mind*. Simon & Schuster, New York.
- [66] Moravec, H. (1998) When will computer hardware match the human brain? *J. Evol. Technol.* 1, 1-8.
- [67] Mountcastle, V.B. (2005) *The sensory hand. Neural mechanisms of somatic sensation*. Harvard University Press, Cambridge, MA.
- [68] Nieoullon A, Coquerel A. (2003) Dopamine: a key regulator to adapt action, emotion, motivation and cognition. *Curr. Opinion in Neurol. Suppl* 2, S3-9.
- [69] Pauling, L. (1940) A theory of the structure and process of formation of antibodies. *J. Amer. Chem. Soc.* 62, 2643-2657.
- [70] Pennisi, E. (2006) Brain evolution on the far side. *Science* 314, 244-245.
- [71] Piattelli-Palmarini, M. (1989) Evolution, selection and cognition: From „learning“ to parameter setting in biology and in the study of language. *Cognition* 31, 1-44.
- [72] Pinsk, M.A., Kastner, S. (2007) Unconscious networking. *Nature* 447, 46-47.
- [73] Purich, D.L. (2001) Enzyme catalysis: a new definition accounting for noncovalent substrate- and product-like states. *Trends in Biochem. Sci.* 2, 417-421.
- [74] Quartz, S.R. (1999) The constructivist brain. *Trends in Cogn. Sci.* 3, 48-57.
- [75] Raichle, M.E. (2006) The brain's dark energy. *Science* 314, 1249-1250.
- [76] Raichle, M.E., Gusnard, D.A. (2002) Appraising the brain's energy budget. *Proc. Natl. Acad. Sci.* 99, 10237-10239.
- [77] Raichle, M.E., Gusnard, D.A. (2005) Intrinsic brain activity sets the stage for expression of motivated behavior. *J. Compar. Neurol.* 493, 167-176.
- [78] Rose, S.P.R. (2005) Human agency in the neurocentric age. *EMBO Rep.* 6, 1001-1005.

- [79] Roth, G., Dicke, U. (2005) Evolution of the brain and intelligence. *Trends in Cogn. Sci.* 9, 250-257.
- [80] Scharrer, B. (1976) Neurosecretion - comparative and evolutionary aspects, str. 125-137. V: Corner, M. A., Swaab, D. F. (eds.) *Perspective in brain research.* Elsevier, Amsterdam.
- [81] Sejnowski, T. (2003) The computational self. *Ann. N.Y. Acad. Sci.* 1001, 262-271.
- [82] Shulman, R.G., Rothman, D.L. (1998) Interpreting functional imaging studies in terms of neurotransmitter cycling. *Proc. Natl. Acad. Sci. U.S.* 95, 11993-11998.
- [83] Shulman, R.G., Rothman, D.L., Behar, K.L., Hyder, F. (2004) Energetic basis of brain activity: implications for neuroimaging. *Trends in Neurosci.* 27, 489-495.
- [84] Stoka, A.M. (1999) Phylogeny and evolution of chemical communication: an endocrine approach. *J. Mol. Endocrinol.* 22, 207-225
- [85] Syková, E. (2005) Glia and volume transmission during physiological and pathological states. *J. Neural Transm.* 112-147.
- [86] Tarrant, A.M. (2005) Endocrine-like signaling in Cnidarians: Current understanding and implications for ecophysiology. *Integrative and Compar. Biol.* 45, 201-214. *biology* 45, 201-214.
- [87] Thorpe, S.J., Fabre-Thorpe, M. (2001) Seeking categories in the brain. *Science* 291, 260-263.
- [88] Trevathan, W.R. (1987) *Human Birth.* Aldine de Gruyter, New York.
- [89] Tsodyks, M., Kenet, T., Grinvald, A., Arieli, A. (1999) Linking spontaneous activity of single cortical neurons and the underlying functional architecture. *Science* 286, 1943-1946.
- [90] Van Gelder, T. J. (1998) The dynamical hypothesis in cognitive science. *Behav. Brain Sci.* 21, 1-14.
- [91] Vincent, J.-D. (1986) *Biologie des passions.* Odile Jacob, Paris.
- [92] Vincent, J.L., Patel, G.H., Fox, M.D., Snyder, A.Z., Baker, J.T., Van Essen, D.C., Zempler, J.M., Snyder, L.H., Corbetta, M., Raichle, M.E. (2007) Intrinsic functional architecture in the anaesthetized monkey brain. *Nature* 447, 83-86.
- [93] Von Neumann, J. (1958) *The computer and the brain.* Yale University Press, New Haven, CT.
- [94] Walker, A., Shipman, P. (1996) *The wisdom of the bones.* Knopf, New York.
- [95] Wiedermann, J. (2004) Spojení samoorganizace s výpočty: minimální život v moři umělých molekul. V: Kvasnička, V., Kelemen, J. (eds.) *Kognice a umělý život,* Slezská univerzita, Opava.
- [96] Wiesel, T.N. (1982) Postnatal development of the visual cortex and the influence of environment. *Nature* 299, 583-591.
- [97] Wilson, E.O. (1999) *Konsilience.* Lidové noviny, Praha.
- [98] Wilson, F.R. (1999) *The hand. How Its use shapes the brain, language, and human culture.* Random House, New York.
- [99] Wimsatt, W. C. (1986) Developmental constraints, generative entrenchment, and the innate-acquired distinction. V: Bechtel, P.W. (ed.) *Integrating Scientific Disciplines,* str. 185-208. Martinus-Nijhoff, Dordrecht.
- [100] Wright, S. (1930) Review of *The genetical theory of natural selection,* by R.A. Fisher. *J. Hered.* 21, 349-356. Pretlačené v: Provine, W. B. (ed.) 1986, *Sewall Wright, Evolution, Selected Papers,* University of Chicago Press, Chicago.
- [101] Zoli, M., Torri, C., Ferrari, R., Jansson, A., Zini, I., Fuxe, K., Agnati, L.F. (1998) The emergence of the volume transmission concept. *Brain Res. Rev.* 26, 136-147.

Kripkeho sémantický model možných svetov

Vladimír Kvasnička¹

Abstrakt. Cieľom tejto práce je popis vlastností a histórie Kripkeovho sémantického modelu modálnej logiky, ktorý zohral výnimočnú úlohu pri rozvoji moderných neklasických logík.

1 Úvodné poznámky

Americký logik a filozof Saul Kripke patrí medzi najvýznamnejšie postavy americkej vedy z 2. polovice minulého storočia. Jeho najdôležitejší teoretický príspevok k rozvoju vedy je jeho *model* sémantickej interpretácie modálnych a im príbuzných logík, ktorý vypracoval v sérii prác [13–20] na rozhraní 50. a 60. rokov minulého storočia. Možno konštatovať, že tieto publikácie stimulovali prudký rozvoj modálnych a im príbuzných logík, ktoré v súčasnosti patria do “hard-core” matematickej logiky s rozsiahlymi aplikáciami nielen v samotnej matematickej logike ale aj v mnohých oblastiach informatiky a umelej inteligencie. Naším cieľom je prezentovať elementárny úvod do Kripkeho sémantického modelu a tým pripomenúť blížiac sa 50. výročie vytlačenia jeho prvej práce v časopise *Journal of Symbolic Logic* [13], ktorú poslal do redakcie ako 19 ročný.

2 Syntax a sémantika výrokovej logiky

Jeden z prvých formálnych systémov v ľudskej histórii bol Euklidov axiomatický systém geometrie, ktorý sa stal vzorom pre rozvíjanie ďalších systémov hlavne v rámci matematiky, informatiky a logiky. V teórii formálnych systémov sa obvykle od seba striktné separujú tieto dva aspekty:

- (1) *syntax* – spôsob konštrukcie jazyka (formúl) teórie,
- (2) *sémantika* – významová interpretácia jazyka (jednotlivých formúl).

Špecifikácia jazyka výrokovej logiky (syntax)

Jazyk L_0 výrokovej logiky je tvorený formulami výrokovej logiky, ktoré sú definované pomocou množiny atomických výrokových premenných $P = \{p, q, \dots, p', q', \dots\}$ a množiny logických spojok $\{\wedge, \vee, \Rightarrow, \equiv, \neg\}$.

Formuly výrokovej logiky, ktoré tvoria jazyk L_0 , sú rekurentne definované ako minimálna množina, ktorá vyhovuje týmto vlastnostiam

¹ FIIT STU, Bratislava, email: kvasnicka@fiit.stuba.sk

- (1) $P \subseteq L_0$,
- (2) ak $(\varphi \in)$, potom $(\neg\varphi) \in L_0$,
- (3) ak $(\varphi, \psi \in L_0)$, potom $(\varphi \wedge \psi), (\varphi \vee \psi), (\varphi \Rightarrow \psi), (\varphi \equiv \psi) \in L_0$.

Špecifikácia významu výrokovej logiky (sémantika)

Ďalší pojem dôležitý pre výrovkovú logiku je *sémantika*. Pojem pochádza z teórie prirodzených jazykov, kde sémantika špecifikuje význam danej vety (ktorá ma tiež aj svoj syntax). Vo výrokovej logike, ktorá sa zaoberá len pravdivosťnými hodnotami premenných a ich formúl, *sémantika* nie je veľmi bohatá. Sémantika výrokovej formuly je špecifikovaná tabuľkou pravdivosťných hodnôt formuly pre rôzne hodnoty jej výrokov. Formuly (elementy jazyka L_0 výrokovej logiky) majú pravdivosťný význam, ktorý je špecifikovaný takto (symbol '1' reprezentuje 'pravdivý' a symbol '0' reprezentuje 'nepravdivý'):

- (1) **Konjunkcia** $\alpha \wedge \beta$ je pravdivá vtedy a len vtedy (vtt) ak obe jej komponenty sú pravdivé, v opačnom prípade je nepravdivá
 $val(\alpha \wedge \beta) = 1$ vtt $val(\alpha) = 1$ a $val(\beta) = 1$
 $val(\alpha \wedge \beta) = 0$ vtt $val(\alpha) = 0$ alebo $val(\beta) = 0$
- (2) **Disjunkcia** $\alpha \vee \beta$ je pravdivá vtt, ak aspoň jedna jej komponenta je pravdivá, v opačnom prípade je nepravdivá
 $val(\alpha \vee \beta) = 1$ vtt $val(\alpha) = 1$ alebo $val(\beta) = 1$
 $val(\alpha \vee \beta) = 0$ vtt $val(\alpha) = 0$ a $val(\beta) = 0$
- (3) **Implikácia** $\alpha \Rightarrow \beta$ je pravdivá vtt, ak prvá jej komponenta (α) je nepravdivá alebo druhá komponenta (β) je pravdivá, v opačnom prípade (prvá komponenta je pravdivá a druhá komponenta je nepravdivá) je nepravdivá
 $val(\alpha \Rightarrow \beta) = 1$ vtt $val(\alpha) = 0$ alebo $val(\beta) = 1$
 $val(\alpha \Rightarrow \beta) = 0$ vtt $val(\alpha) = 1$ a $val(\beta) = 0$
- (4) **Ekvivalencia** $\alpha \equiv \beta$ je pravdivá vtt, ak obe komponenty majú rovnakú pravdivosťnú hodnotu, v opačnom prípade je nepravdivá
 $val(\alpha \equiv \beta) = 1$ vtt $val(\alpha) = val(\beta)$
 $val(\alpha \equiv \beta) = 0$ vtt $val(\alpha) \neq val(\beta)$
- (5) **Negácia** $\neg\alpha$ je pravdivá vtt, ak jej komponenta je nepravdivá, v opačnom prípade je pravdivá
 $val(\neg\alpha) = 1$ vtt $val(\alpha) = 0$
 $val(\neg\alpha) = 0$ vtt $val(\alpha) = 1$

Formula φ sa nazýva *tautológia* (alebo tautologickým dôsledkom, čo vyjadríme $\models\varphi$), ak pre každú interpretáciu τ pravdivosťných hodnôt jej atomických premenných je formula pravdivá

$$(\models\varphi) =_{def} \forall(\tau)(val_{\tau}(\varphi) = 1)$$

V opačnom prípade, ak pre každú interpretáciu τ platí $val_{\tau}(\varphi)=0$, formula sa nazýva *kontradikcia*. Ak existuje aspoň jedna interpretácia τ taká, že $val_{\tau}(\varphi)=1$, potom formula φ je *splniteľná* (to znamená, že tautológia je špeciálny prípad splniteľnosti). Môžeme teda povedať, že všetky formuly, ktoré nie sú kontradikcie sú splniteľné a tautológie sú také splniteľné formuly, ktoré sú pre všetky možné interpretácie τ pravdivé.

Tautológie majú vo výrokovej logike mimoriadne postavenie *zákonov* logiky, tieto formule sú vždy pravdivé pre ľubovoľné pravdivostné hodnoty premenných. Niektoré tautológie sa často používajú nielen v samotnej výrokovej logike, ale aj v bežnom usudzovaní a sú obvykle označované aj vlastným menom. Väčšinou ide o tautológie tvaru ekvivalencie, ktoré umožňujú nahradzovať jedny formuly inými bez straty vlastností ich tautologickosti.

Veta (tautológia, teorém, zákon, argument, ...) je výrok o ktorom môže byť ukázané, že je pravdivý. V tejto súvislosti hovoríme o *dôkaz*e vety, ktorý spočíva v postupnosti jednotlivých „medzikrokov“, ktoré sú odvodené buď z množiny jednoduchých postulátov – zákonov, nazývaných *axiómy*, alebo z predchádzajúcich viet (pomocných viet, často nazývaných lemy) danej postupnosti. Komplikované dôkazy sú obvykle jasnejšie formulované, keď ich dôkaz je rozdelený na jednotlivé medzikroky, ktoré sú formulované ako samostatné vety. Tieto medzikroky - vety v postupnosti sú vytvárané pomocou *pravidiel odvodzovania* (*pravidiel usudzovania*), ktoré z niekoľkých pravdivých tvrdení - argumentov vytvorí nové pravdivé tvrdenie - argument. Hovoríme, že formula φ logicky vyplýva (čo zapisujeme $\vdash\varphi$) z predpokladov – axióm práve vtedy, ak existuje postupnosť formúl, z ktorých každá je buď východiskovou axiómou alebo bola v predchádzajúcom kroku odvodená z predchádzajúcich formúl pomocou pravidla modus ponens.

V predchádzajúcej časti tejto kapitoly boli formulované dva nezávislé pojmy:

- (1) V syntaktickom prístupe pojem logického dôsledku formuly ψ z množiny axióm, čo formálne zapisujeme $\vdash\psi$.
- (2) V sémantickom prístupe pojem tautologického dôsledku formuly ψ , čo formálne zapisujeme $\models\psi$.

Aj keď sú tieto pojmy definované nezávislým spôsobom (syntaktickým a sémantickým), existuje medzi nimi tesná väzba, podľa viet o úplnosti a korektnosti výrokovej logiky platí:

Pre ľubovoľnú formulu ψ vzťah $\vdash\psi$ platí práve vtedy, ak $\models\psi$
 $(\vdash\psi)\equiv(\models\psi)$

Z tejto vety vyplýva ekvivalentnosť syntaktického a sémantického prístupu, je úplne jedno, či použijeme syntaktickú metódu na stanovenie relácie $\vdash\psi$ alebo sémantickú metódu na stanovenie relácie $\models\psi$. Veta o ekvivalentnosti úplnosti a korektnosti patrí medzi základné výsledky výrokovej logiky. Na jej základe sme oprávnení dokázaťnosť nejakej formuly preverovať tým, že dokážeme jej tautologičnosť, ktorá je definovaná prostredníctvom sémantického pojmu pravdivostného ohodnotenia, napr. pomocou tabuľkovej metódy. Syntaktický pojem dokázaťnosti splyva so sémantickým pojmom tautologičnosti, čo je jedinečná vlastnosť výrokovej logiky a ojedinelá vlastnosť formálnych systémov, kde obvykle existuje zreteľná demarkačná čiara medzi syntaxou a sémantikou daného systému. Na záver môžeme teda konštatovať, že výrovková logika je

- **korektná** (ak každá dokázaná formula z axióm je tautológia),
- **nerozporná** (ak zo systému axióm súčasne logicky nevyplývajú formuly φ a $\neg\varphi$),
- **úplná** (ak každá tautológia je dokázaťná z axióm) a
- **rozhodnutel'ná** (existuje jednoduchý algoritmus, pomocou ktorého sme schopní rozhodnúť či pre dané pravdivostné hodnoty premenných je formula pravdivá alebo nie).

3 Modálna logika [3,21,27]

Použijeme jednoduchý prístup ako zaviesť modálnu logiku pomocou výrokovej logiky. Množina logických spojok výrokovej logiky, $\{\wedge, \vee, \equiv, \neg\}$, je rozšírená o ďalšie dve unárne spojky \Box a \Diamond . Formula modálnej logiky $\Diamond\varphi$ znamená 'nutne platí φ ', analogická formula $\Box\varphi$ znamená 'možno platí φ '. Tieto dve spojky nie sú nezávislé, navzájom sú zviazané reláciou $\Box\varphi \equiv \neg\Diamond\neg\varphi$, alebo $\Diamond\varphi \equiv \neg\Box\neg\varphi$. Pôvodne boli tieto dve logické spojky \Box a \Diamond chápané ako validita resp. splniteľnosť. Analogické spojky boli neskoršie použité pri formulácii kauzálnej, deontickej, epistemickej a veľmi populárnej temporálnej logiky.

Ako jednoduché rozšírenie klasickej výrokovej logiky pomocou modálnych spojok boli študované rôzne typy modálnych logík americkým logikom C. I. Lewisom, ktorý je pokladaný za zakladateľa modernej modálnej logiky [22]. Žiaľ, jeho spôsob prezentácie modálnych logík bol veľmi ťažkopádny, aj keď zo súčasného pohľadu väčšina jeho výsledkov sa ukázala korektnými. Zásluhou Kurta Gödela [5] tento systém bol prevedený do súčasného prehľadného hierarchického axiomatického systému, kde zapájaním rôznych axióm dostávame rôzne typy modálnych logík. V tomto prístupe veľmi dôležitú úlohu

hrá nové pravidlo odvodzovania (okrem štandardného pravidla modus ponens) nazývané pravidlo nutnosti² $\varphi/\Box\varphi$, ktoré umožňuje inferenciu $\Box\varphi$ z východiskového predpokladu φ . Najdôležitejšie axiómy tohto systému sú

- (Ax₀) $\Box(\varphi \Rightarrow \psi) \Rightarrow (\Box\varphi \Rightarrow \Box\psi)$
- (Ax₁) $\Box\varphi \Rightarrow \varphi$
- (Ax₂) $\Box\varphi \Rightarrow \Box\Box\varphi$
- (Ax₃) $\Box\varphi \Rightarrow \Box\neg\Box\neg\varphi$

Najdôležitejšie systémy vždy obsahujú formulu (Ax₀); systém, ktorý obsahuje len túto axiómu a pravidlo nutnosti sa nazýva **K** na počesť Saula Kripkeho, pretože mal mimoriadne postavenie pri vzniku jeho teórie modelu modálnej logiky. Ostatné najdôležitejšie systémy modálnej logiky vznikli zahrnutím týchto axiém

- K:** (Ax₀)
- T:** (Ax₀) + (Ax₁)
- S4** (Ax₀) + (Ax₁) + (Ax₂)
- B:** (Ax₀) + (Ax₁) + (Ax₃)
- S5:** (Ax₀) + (Ax₁) + (Ax₂) + (Ax₃)

V počiatočných etapách (hlavne 40. a 50. roky minulého storočia) úloha konštrukcie modelu v modálnej logike bola realizovaná zvlášť pre daný typ tejto logiky špecifikovanej vyššie pomocou axiomatického systému (Ax₀ – Ax₃) modálnej logiky. Všeobecnou snahou boli pokusy navrhnuť spoločný model pre modálne logiky rôzneho typu, ktorý by prechádzal na iný typ modelu len jednoduchou zmenou základných „variabilných parametrov“ modelu. Jeden zo základných problémov, ktorý pri týchto pokusoch musel byť vyriešený je zovšeobecnenie štandardného modelu výrokovej logiky tak, aby prirodzene obsahoval aj model *M* danej modálnej logiky rozšírením unárnych spojok o dve nové spojky nutnosti a možnosti \Box resp \Diamond .

$\Box\varphi$ je pravdivé v <i>M</i>	vtt	φ je nutne pravdivé v <i>M</i>
$\Diamond\varphi$ je pravdivé v <i>M</i>	vtt	φ je možné pravdivé v <i>M</i>

Formuly modálnej logiky (reťazce písmen reprezentujúcich premenné, konštant **0** a **1** priradených významu pravdivosti resp. nepravdivosti, a logických spojok) sú formálne (matematické objek-

² Toto pravidlo v rámci sémantickej interpretácie môžeme jednoducho interpretovať tak, že ak formula φ je tautológia, potom aj $\Box\varphi$ musí byť tautológia.

ty) a ako také, ich sémantický pravdivostný význam v rámci modelu M , je taktiež len určitá formálna (matematická) skutočnosť. Pretože matematické skutočnosti sú vždy nutné, potom pravdivosť/nepravdivosť formuly φ v rámci modelu M nám musí implikovať pravdivosť/nepravdivosť $\Box\varphi$, takže $\varphi \Rightarrow \Box\varphi$ je platná formula. Z týchto jednoduchých úvah nám vyplýva, že ak φ je pravdivá formula pre model M , potom musí byť pravdivá aj formula $\Box\varphi$, bez ohľadu, či táto skutočnosť v použitom modeli M platí alebo nie, takže $\varphi \Rightarrow \Box\varphi$ je platná formula. To znamená, aby sme sa vyhli týmto „paradoxom“ jednoduchého modelu M pre vyhodnotenie pravdivosti formuly $\Box\varphi$, musíme používať zložitejšie teoretické prístupy ku konštrukcii modelu ako je tento „triviálne jednoduchý“ použitý model.

4 Kripkeho sémantický model pre modálnu logiku

Kripkeho model modálnej logiky je podstatne zložitejší ako ten, ktorý bol použitý v závere predchádzajúcej kapitoly. Pravdivostné ohodnotenie danej formuly modálnej logiky bude založené na rekurzii podľa zložitosti formuly, pričom môžeme prechádzať z daného „sveta“, kde vyhodnotenie realizujeme, do iného „alternatívneho sveta“, kde využívame pravdivostné ohodnotenie jej podformúl. Snáď na tejto úvodnej úrovni môžeme použiť vysvetľujúcu analógiu so situáciou, keď chceme riešiť pravdivosť výroku $\Box\varphi$, kde $\varphi =$ 'vlak do Žiliny odchádza o 14.37 hod'. Ak nám všetci kolegovia v práci (reprezentujúci Kripkeho alternatívne svety) vyhodnotia výrok φ ako pravdivý, potom môžeme povedať, že formula $\Box\varphi$ je pravdivá, alebo výrok 'vlak do Žiliny odchádza o 14.37 hod' je nutne pravdivý výrok. Alternatívne, ak nám výrok φ vyhodnotí aspoň jeden kolega ako pravdivý, potom hovoríme, že výrok 'vlak do Žiliny odchádza o 14.37 hod' je možné pravdivý, čo označíme ako $\Diamond\varphi$.

Kripkeho model $M = (W, R, w_0, val)$ je usporiadaná štvorica. Symbol $W = \{w, w', \dots\}$ je neprázdna množina svetov, symbol $R \subseteq W \times W$ je binárna relácia, $w_0 \in W$ je odlíšený svet pre daný model, a val je zobrazenie, ktoré pravdivostne ohodnocuje výrokové premenné z $P = \{p, q, \dots, p', q', \dots\}$ pre daný svet $w \in W$, $val: W \times P \rightarrow \{0, 1\}$. Symbol $val(w, p) = (1, 0)$ znamená, že premenná $p \in P$ je vo svete $w \in W$ pravdivá (nepravdivá). Binárna relácia R špecifikuje tzv. dostupné svety, ak $(w, w') \in R$, potom hovoríme, že svet w' je dostupný zo sveta w . Binárna relácia R môže byť formálne nahradená podmnožinami $\Gamma(w) \subseteq W$, kde $\Gamma(w) = \{w' \in W \mid (w, w') \in R\}$ obsahuje svety, ktoré sú dostupné zo sveta w . Odlíšený svet w_0 špecifikuje použitý model M pre daný svet w_0 . Zobrazenie val je zovšeobecnené pre neatomické formuly φ a ψ takto:

(1)	$\neg\varphi$ je pravdivé pre svet w_0 v modeli M	vtt	ak φ je nepravdivé pre w_0 a M
(2)	$\varphi \wedge \psi$ je pravdivé pre svet w_0 v modeli M	vtt	ak φ je pravdivé pre w_0 v M a ψ je pravdivé pre w_0 v M
(3)	$\varphi \vee \psi$ je pravdivé pre svet w_0 v modeli M	vtt	ak φ je pravdivé pre w_0 v M alebo ψ je pravdivé pre w_0 v M
(4)	$\varphi \Rightarrow \psi$ je pravdivé pre svet w_0 v modeli M	vtt	ak φ je nepravdivé pre w_0 v M alebo ψ je pravdivé pre w_0 v M
(5)	$\Box\varphi$ je pravdivé pre svet w_0 v modeli M	vtt	ak φ je pravdivé pre každý $w' \in \Gamma(w)$ v M
(6)	$\Diamond\varphi$ je pravdivé pre svet w_0 v modeli M	vtt	ak φ je pravdivé pre niektoré $w' \in \Gamma(w)$ v M

V tejto tabuľke riadok (6) je redundantný, danú podmienku môžeme odvodiť z ekvivalencie $\Diamond\varphi \equiv \neg\Box\neg\varphi$. Takto naformulovaný Kripkeho model často nazývame aj ako **Kripkeho sémantika** pre výrokovú modálnu logiku. Kripke pri konštrukcii tohto modelu bol stimulovaný Leibnizom, ktorý spomína možné svety, ktoré sú obsiahnuté v myslí Boha, ktorý však stvoril z týchto alternatívne možných svetov ten najlepší svet. Musíme však poznamenať, že táto „pitoreskná“ terminológia Kripkeho sémantického modelu spôsobila mnohé nedorozumenia a hádky pri diskutovaní filozofických základov modálnej logiky. Podobná situácia nastala aj v informatike pri formulovaní všeobecných základov kvantového počítania [6], kde sa taktiež používa koncepcia možných alebo alternatívnych svetov, pomocou ktorej sa hľadá jedno z možných riešení daného problému.

Na záver tejto kapitoly zavedieme štandardnú notáciu modálnej logiky, ktorá bude vyjadrovať skutočnosť, že formuly $\Box\varphi$ a $\Diamond\varphi$ sú pravdivé vo svete w v modeli M , $w \models \Box\varphi$ resp. $w \models \Diamond\varphi$; nepravdivosť týchto formúl sa vyjadruje $w \not\models \Box\varphi$ resp. $w \not\models \Diamond\varphi$. Potom podmienky (5) a (6) z tabuľky uvedenej vyššie môžeme prezentovať takto

$$(w \models \Box\varphi) =_{def} \begin{cases} \bigwedge_{w' \in \Gamma(w)} (w' \models \varphi) & (pre\Gamma(w) \neq \emptyset) \\ \mathbf{1} & (pre\Gamma(w) = \emptyset) \end{cases}$$

$$(w \models \Diamond\varphi) =_{def} \begin{cases} \bigvee_{w' \in \Gamma(w)} (w' \models \varphi) & (pre\Gamma(w) \neq \emptyset) \\ \mathbf{0} & (pre\Gamma(w) = \emptyset) \end{cases}$$

kde bolo vykonané aj zovšeobecnenie formúl pre prípad, že podmnožina $\Gamma(w)$ obsahujúca dostupné svety zo sveta w je prázdna.

5 Korektnosť a úplnosť modálnej logiky

Kripke dokázal, že systém **K** modálnej logiky je korektný a úplný. To znamená, že ak sa nám podarí dokázať pomocou axiomatického systému a niekoľkých pravidiel inferencie logickú validitu (logické vyplývanie) nejakej formuly, potom táto skutočnosť je ekvivalentná s tým, že táto formula je v každom Kripkeho modeli pravdivá (t. j. je tautológia). Táto vlastnosť ekvivalencie medzi syntaktickým a sémantickým odvodzovaním formúl patrí vo výrokovej logike medzi jej fundamentálne teoretické výsledky, ako vidíme, podobná vlastnosť platí aj vo výrokovej modálnej logike **K**.

V iných typoch modálnych logík (napr. **T**) platí taktiež podobná vlastnosť korektnosti a úplnosti, len v týchto prípadoch musíme už bližšie špecifikovať vlastnosti relácie R , ktorá je integrálnou súčasťou Kripkeho sémantiky (Pre **K** modálnu logiku relácia R nemusí byť bližšie špecifikovaná). V **T** modálnej logike formuly – teóremy, ktoré sú odvodené z axiomatického systému sú pravdivé v Kripkeho sémantike len vtedy, ak relácia R je reflexívna, $(\forall w)((w, w) \in R)$ alebo $(\forall w)(w \in \Gamma(w))$. To znamená, že modálna logika **T** je korektná práve vtedy, ak každý teorém vyplývajúci z axiomatického systému je pravdivý v Kripkeho sémantike s reflexívnou reláciou dostupnosti R , a naopak, modálna logika **T** je úplná ak každá tautológia pre sémantické modely s reflexívnou reláciou R je odvoditeľná v rámci axiomatického systému **T**. Vzťah medzi schémou axióm $(Ax_0) + (Ax_1)$ (o ktorý je rozšírený Hilbertov systém axióm) a reflexívnou Kripkeho sémantikou je intuitívne plne pochopiteľný. Pravdivosť formuly $\Box\phi$ vo svete w implikuje, že formula ϕ je pravdivá v každom svete z $\Gamma(w)$, teda aj vo svete w , pretože $w \in \Gamma(w)$, čo bolo potrebné dokázať.

Ostatné schémy výbery axióm sú sémanticky interpretované v rámci Kripkeho modelu pomocou iných podmienok, ktoré sú požadované od binárnej relácie R . Tak napríklad axióm (Ax_2) je na sémantickej úrovni interpretovaný reflexívnou reláciou R (t. j. $w \in \Gamma(w) \wedge w'' \in \Gamma(w') \Rightarrow w'' \in \Gamma(w)$). Predpokladajme pravdivosť $\Box\phi$ vo svete w , potom formula ϕ je pravdivá v každom svete z $\Gamma(w)$. Podmienka tranzitivity nám zaručuje, že ak platí $w' \in \Gamma(w)$ a pre svet w'' platí $w'' \in \Gamma(w')$, potom aj $w'' \in \Gamma(w)$. To znamená, že formula $\Box\phi$ je pravdivá aj vo svete w'' , z čoho vyplýva, že $\Box\Box\phi$ je pravdivá taktiež vo svete w . Potom $\Box\phi \Rightarrow \Box\Box\phi$.

Podobným spôsobom môžeme dokázať, že axióm (Ax_3) korešponduje s podmienkou symetričnosti relácie R . Kripke ukázal, že mo-

dálna logika **S4** je korektná a úplná pre reflexívny a tranzitívny model Kripkeho sémantiky a taktiež ukázal, že modálna logika **B** má taktiež tieto dve vlastnosti pre sémantický model s reláciou R , ktorá je reflexívna, symetrická a tranzitívna (t. j. R je relácia ekvivalentnosti). Na záver treba poznamenať, že Kripkeho sémantický model umožnil jednoduchú klasifikáciu rôznych typov modálnej logiky pomocou podmienok kladených na reláciu R . Táto skutočnosť veľmi plodne stimulovala ďalší rozvoj modálnej logiky pre potreby filozofie, umeljej inteligencie a informatiky, v súčasnosti patrí medzi najbúrlivejšie sa rozvíjajúce oblasti logiky s veľmi veľkým aplikačným potenciálom v rôznych oblastiach ľudského poznania.

6 Kritika Kripkeho sémantického modelu

V tejto časti práce pokladáme za potrebné upozorniť čitateľa na dva aspekty Kripkeho sémantického modelu, ktoré často bývajú predmetom jeho kritiky:

- (1) Ohodnotenie *val* je integrálnou súčasťou Kripkeho modelu, pričom toto ohodnotenie je špecifikované len veľmi všeobecne. Fínsky logik Jaakko Hintikka píše [9]: *Čo pokladáme za potrebné zdôrazniť pri špecifikácii podmienky, že výroková premenná p je nutne pravdivá vo svete w_0 je skutočnosť, že pravdivosť premennej p musí byť špecifikovaná v každom logicky možnom svete. Žiaľ, táto požiadavka nie je splnená v Kripkeho modeli, aby tieto svety patrili do množín dostupných svetov zo sveta w_0 . Hintikka vo svojej kritike Kripkeho sémantiky nešpecifikuje čo sú „logicky možné svety“, ako sa tieto svety líšia od dostupných svetov, ktoré sú špecifikované množinou $\Gamma(w_0)$. Nie je jednoduché pozornému čitateľovi odpovedať na túto Hintikkovu pripomienku. Treba poznamenať, že pre daný Kripkeho model M zobrazenie $val: W \times P \rightarrow \{0, 1\}$ môže byť ľubovoľné a je integrálnou súčasťou daného modelu a nemá zmysel podrobnejšie diskutovať, prečo zvolené zobrazenie je práve také, aké je. Pri definícii napr. tautológie formuly $\Box\varphi$ táto konkrétnosť zobrazenia *val* daného modelu M je „vystredovaná“ tým, že študovaná formula musí byť pravdivá vo všetkých Kripkeho modeloch*

$$(\models \Box\varphi) =_{def} (\forall M)(M \models \varphi) =_{def} \bigwedge_M (M \models \varphi)$$

t. j. koncepcia tautologičnosti nie je priamo vzťahnutá k danému modelu M .

- (2) Ďalšia vážna kritika Kripkeho sémantického modelu je použitá „pitoreskná“ terminológia možných svetov, ktorá bola odštartovaná prehnanou a niekedy až smiešnou predstavivosťou niektorých filozofov a logikov. Tak napríklad D. Lewis [23] chápe túto terminológiu s maximálnou vážnosťou a diskutuje mnohé jej implikácie k rozšírenému chápaniu nášho sveta ako alternatívneho a jedného z možných svetov. Treba poznamenať, že podobná terminológia a pohľad na náš svet je v súčasnosti veľmi populárna v kvantovej fyzike, kde možnosť merania ľubovoľného stavu kvantového systému je vysvetľovaná alternatívne pomocou *hypotézy mnohých svetov*. Vo fyzike táto hypotéza sa pokúša interpretovať nedeterministické procesy (ako napr. meranie), pričom tzv. kolaps vlnovej funkcie je interpretovaný pomocou kvantovej superpozície stavov nekonečne mnohých a identických „paralelných vesmírov“. Hypotéza mnohých svetov sa nezaobera problémom modalít výrokov, ktoré sú riešené pomocou modelu možných svetov. Hlavný rozdiel medzi týmito dvoma koncepciami bez ohľadu na ich pôvod a použitie spočíva v týchto dvoch bodoch: (a) stavy kvantovo-teoretických svetov sú entanglované kvantovo-mechanicky, zatiaľ čo entanglovanie možných svetov sa považuje za bezvýznamné, (b) vo filozofických teóriách možných svetov tieto sú logicky ale nie fyzikálne možné, pričom kvantovo-teoretické svety sú fyzikálne možné. Zaujímavá je aplikácia teórie možných svetov v literárnej vede [26], kde je použitá k interpretácii diel sci-fi a fantasy. Tento teoretický prístup poskytuje vhodný jazyk a konceptuálne rámce pre popis takýchto svetov. Avšak je potrebné podotknúť, že literárne svety sú špecifické typy možných svetov, podstatne odlišné od svetov používaných v logike. Táto skutočnosť je spôsobená tým, že literárne texty sú „obývané“ bytosťami s vlastnými modalitami, literárne svety sú obdarené vlastnou autonómiou podobným spôsobom, ako náš skutočný svet.

Na záver tejto kritiky Kripkeho sémantického modelu upozorníme na nejednoznačnosť slova „sémantika“. Tento termín sa občas používa ako synonymum pre „teória modelu“ s taktiež sa používa ako označenie pre teóriu významu (theory of meaning). V tejto práci sa budeme držať konvencie, že keď budeme hovoriť o sémantickom modeli, vždy použijeme adjektívum „Kripkeho“. Sémantický model ako taký sa môže používať v teoretickej lingvistiky, kde má úplne iný význam ako v modálnej logike, týmto jednoduchým spôsobom odlíšime náš sémantický model od ostatných podobných modelov, ktoré sa hlavne využívajú v jazykovede.

7 Poznámky k histórii sémantického modelu modálnej logiky³

Ako vyplýva z dejín vedy, skoro žiadny objav vo vede nebol vykonaný ako sólo akcia vynikajúceho jedinca. Hovoríme, že doba bola „tehotná“ na ten-ktorý objav, často skoro súčasne základné idey objavu boli formulované na rôznych miestach rôznymi autormi; pripomeňme len dve dvojice: Charlesa Darwina a Alfreda Wallaceho a taktiež Alberta Einsteina a Henri Poincarého. Kripke formuloval základné myšlienky svojho sémantického modelu modálnej logiky ako gymnazista, pričom vychádzal z už známych výsledkov, ktoré vznikli keď bol žiakom základnej školy. Medzi jeho prvé fundamentálne výsledky patrí dôkaz korektnosti a úplnosti modálnej logiky **S4**. K dôkazu týchto vlastností použil techniku sémantických tabiel, ktorá bol naformulovaná v r. 1955 holandským logikom Bethom [1], čiže patrila medzi čerstvé výsledky matematickej logiky.

Pripomenieme si niektoré staršie výsledky o sémantike modálnej logiky. Predovšetkým ide o dve práce Tarského a jeho študentov McKinseyho a Jónssona. Prvá práca [25] z r. 1948 sa zaoberala konštrukciou algebraického modelu modálnej logiky pomocou špeciálnych Boolových algebier. Druhá práca [11] z r. 1951 bola venovaná štúdiu algebraických štruktúr nazývaných rámce, pričom na relácie sa kládli ohraničujúce podmienky. Treba poznamenať, že aj tieto práce obsahovali spolu to, čo dnes označujeme ako Kripkeho sémantický model. Tarskeho a jeho študentov nenapadlo, že tieto práce úzko súvisia a tvoria vlastne sémantický model modálnej logiky, pričom vlastnosti modelu boli špecifikované požiadavkami kladenými na reláciu z rámca. To znamená, že problém priority pre túto trojicu je neobyčajne zamotaný a doplatili na to, že ich spoločný a hlavný autor A. Tarsky nehľadal vzájomné súvislosti medzi týmito dvoma prácami a nevykonali ten „malý krok“ ku vzniku sémantického modelu modálnej logiky.

V ďalšej časti tejto kapitoly budeme diskutovať dva pokusy o vypracovanie sémantického modelu modálnej logiky, kde boli publikované hypotézy o tom, že medzi modálnymi logikami typu **T**, **S4** a **S5** a podmienkami typu reflexivity, tranzitivity a symetrie existuje úzka súvislosť, avšak neboli publikované dôkazy týchto hypotéz.

Švédsky logik Stig Kanger vo svojej PhD-dizertácii z r. 1957 [12] vypracoval sémantický model modálnej logiky. Aj keď táto dizertácia bola publikovaná spôsobom obvyklým pre dizertácie vo Švédsku, nikdy nebola publikovaná časopiseckou formou. Až v r. 2001 bola

³ Táto kapitola vychádza z práce Goldblatta [4], v ktorej prezentuje rozsiahly kritický náčrt histórie modálnej logiky.

táto práca znovu publikovaná v knižnom zborníku venovanom pamiatke Kangeru [10]. Uvedená práca je extrémne ťažko čitateľná v dôsledku novej a neobvyklej terminológie. Už len nálež podmienok tranzitivity, reflexivity a symetričnosti vyžaduje od čitateľa veľké úsilie, aby tieto štandardné podmienky dekodoval v texte dizertácie.

O mnoho bližšie ku konštrukcii sémantického modelu modálnej logiky bol fínsky logik Jaakko Hintikka [8], ktorý v r. 1963 publikoval prácu venovanú tomuto problému. V porovnaní s prácou Kripkeho, Hintikka je menej jasný, jeho sémantika namiesto množiny možných svetov obsahuje množinu formúl. Je potrebné explicitne poznamenať, že práca neobsahuje dôkaz úplnosti a korektnosti jeho modálnej logiky, čo jeho prácu redukuje na rozšírený abstrakt.

Na záver tejto kapitoly môžeme konštatovať, že prvenstvo Kripkeho pri tvorbe sémantického modelu modálnej logiky nie je založené len na tom, že bol prvý, bez predchodcov, ktorý vytvoril korektný sémantický model bezprostredne aplikovateľný k rôznym typom modálnej logiky alebo, že bol prvý, ktorý poskytol dôkazy korektnosti a úplnosti modálnych logík rôznych typov, ale v tom, že ako prvý poukázal na neobyčajnú vhodnosť jeho sémantického (Kripkovho) modelu tak pre rôzne typy modálnej a intuicionistickej logiky ako aj pre rôzne typy predikátovej modálnej a intuicionistickej logiky. V poslednej časti tejto kapitoly upriamime našu pozornosť práve na tie dve možnosti použitia jeho sémantického modelu na intuicionistickú logiku a predikátovú modálnu logiku.

8 Kripkeho model intuicionistickej logiky

Počiatkom 20. storočia holandský matematik a logik L. E. J. Brouwer položil základy intuicionistickej filozofie matematiky a logiky, ktorej základným krédom je postulát, že neexistujú matematické objekty *a priori* dané, ktoré sú nezávislé na ľudskej aktivite. Do popredia v intuicionistickej filozofii vystupujú konštruktivistické aspekty matematiky, ako prejav ľudských snáh zostrojovať nové formálne štruktúry. Preto v súčasnej literatúre sa často používa termín *konštruktivizmus* ako ekvivalent intuicionistického pohľadu na matematiku a logiku.

Konečným cieľom konštruktivizmu je preformulovať celú matematiku a logiku tak, že ak sa má dokázať existencia nejakého objektu x , ktorý má vlastnosť $P(x)$, potom dôkaz tejto vlastnosti musí byť založený na algoritme, ktorý zabezpečí konštrukciu objektu x s vlastnosťou P . To znamená, že v konštruktivistickej matematike vôbec neexistujú existenčné teorémy (obvykle založené na dôkaze sporom), každý teorém musí mať konštruktívny charakter, t. j. musí obsahovať

predpis ako zostrojil' objekt s požadovanou vlastnosťou. V súčasnej matematike a logike existujú veľké pochybnosti, či tento konštruktivistický cieľ je plne uskutočniteľný. Ukazuje sa, že v mnohých prípadoch sa konštruktívny dôkaz hľadá veľmi zložitými cestami (ak vôbec existuje), preto v štandardnej matematike a logike musia spolu koexistovať tak existenčné ako aj konštruktívne dôkazy jednotlivých vlastností, obmedziť sa len na konštruktívne dôkazy by znamenalo, že výstavba danej matematickej disciplíny by sa stala veľmi ťažkopádnu⁴.

Základné *krédo intuicionizmu* v matematike a logike je špecifikované pomocou týchto štyroch BHK⁵ podmienok požadovaných od konštruktívnych dôkazov:

BHK podmienky.

- konštruktívny dôkaz $p \wedge q$ je tvorený z konštruktívnych dôkazov p a q ,
- konštruktívny dôkaz $p \vee q$ pozostáva z náhodného výberu p alebo q a z konštruktívneho dôkazu zvoleného výroku,
- konštruktívny dôkaz $p \Rightarrow q$ je tvorený takým algoritmom, ktorý každý konštruktívny dôkaz p pretransformuje na konštruktívny dôkaz q ,
- negácia $\neg p$ je chápaná ako implikácia $p \Rightarrow 0$; konštruktívny dôkaz $\neg p$ je tvorený takým algoritmom, ktorý každý konštruktívny dôkaz p pretransformuje na nepravdu (ináč povedané, konštruktívny dôkaz $\neg p$ je tvorený takým postupom, ktorý ukáže, že neexistuje konštruktívny dôkaz p).

Prečo týmto radikálnym spôsobom reinterpretujeme klasickú logiku? Hlavný dôvod k zavedeniu tohto nového netradičného pohľadu na klasickú (aristotelovskú) logiku je požiadavka zachovať čo najviac konštruktivistický (výpočtový) charakter matematiky. Podľa intuicionizmu, výrok p je pravdivý len vtedy, ak existuje jeho konštruktívny dôkaz.

Holandský logik a matematik Arend Heyting v 30. rokoch minulého storočia [7] navrhol axiomatický systém intuicionistickej logiky, ktorý postihuje základné myšlienky intuicionistickej filozofie mate-

⁴ V tejto súvislosti sa často spomína jednoduchý ilustračný príklad o tom, že ak spojitá funkcia f , definovaná na intervale $\langle 0, 1 \rangle$, vyhovuje podmienke $f(0) \cdot f(1) < 0$, potom existuje také číslo $0 < c < 1$, že $f(c) = 0$. Existenčný dôkaz tejto vlastnosti sporom je jednoduchá záležitosť, avšak konštruktívny dôkaz tejto vlastnosti je neobvyčajne komplikovaný a zdĺhavý.

⁵ Akronym mien **Brouwer**, **Heyting** a **Kolmogorov**, matematikov a logikov z prvej polovice 20. storočia, ktorí sa zaslúžili o rozvoj intuicionizmu v matematike a logike.

matiky a logiky. Formuly obsahujú výrokové premenné a štyri logické spojky $\{\wedge, \vee, \Rightarrow, \neg\}$. To znamená, že jazyk výrokovvej logiky a jazyk intuicionistickej logiky je rovnaký, avšak tieto dve logiky sa líšia v tom, či nejaká formula ϕ je zákonom (teorémou) alebo nie je. Platí, že každý zákon intuicionistickej logiky je aj zákonom výrokovvej logiky, inverzia tohto tvrdenia však neplatí. Môžeme teda konštatovať, že tieto logiky majú rovnaké jazyky a množina zákonov intuicionistickej logiky je podmnožinou množiny zákonov výrokovvej logiky.

Pre výrokovú logiku majú BHK podmienky dramatický dopad na skutočnosť, či nejaká formula je zákonom (tautológiou) výrokovvej logiky alebo nie. Použitím BHK podmienok dokážeme, že implikácia $p \Rightarrow q$ nie je prevoditeľná na disjunkciu a negáciu $\neg p \vee q$, t. j. $(p \Rightarrow q) \neq_{def} (\neg p \vee q)$. Konštruktivistický dôkaz implikácie $p \Rightarrow q$ spočíva v tom, že existuje taký algoritmus, ktorý každý dôkaz p pretransformuje na dôkaz q . Alternatívna interpretácia pravej strany $\neg p \vee q$ je taká, že dôkaz $\neg p$ je tvorený algoritmom, ktorý každý dôkaz p pretransformuje na konštruktívny dôkaz kontradikcie. Tento záver však nesúhlasí s prvým záverom, podľa ktorého každý dôkaz p je pretransformovaný na dôkaz q .

V intuicionistickej logike klasické spojky \neg a \Rightarrow majú neklasický význam. Ako už bolo povedané, intuicionisti chápu každý dôkaz negácie $\neg\phi$ ako konštruktívny dôkaz kontradikcie z predpokladu ϕ . Podobným spôsobom interpretujú aj implikáciu, kde každý dôkaz formuly $\phi \Rightarrow \psi$ je chápaný ako konštruktívny dôkaz dôsledku ψ z predpokladu ϕ . Takýto pohľad v intuicionistickej logike na dôkaz formúl $\neg\phi$ a $\phi \Rightarrow \psi$ sa realizuje tak, že formálne prepíšeme formulu intuicionistickej logiky ϕ na pomocnú formulu $\phi^* = \Box\phi$, kde modálna spojka \Box sa chápe ako konštruktívny dôkaz. Potom konštruktívny dôkaz negácie $\neg\phi$ je realizovaný ako $(\neg\phi)^* = \Box\neg\phi$. Transformácia implikácie má tvar $(\phi \Rightarrow \psi)^* = \Box(\phi \Rightarrow \psi^*)$. Pre spojky konjunkcie a disjunkcie platia tieto jednoduché transformácie: $(\phi \wedge \psi)^* = \phi^* \wedge \psi^*$ resp. $(\phi \vee \psi)^* = \phi^* \vee \psi^*$. Tento transformačný postup bol navrhnutý Gödelom [5] v r. 1932, ktorý vyslovil domnienku, že formula ϕ patrí do intuicionistickej logiky práve vtedy, ak formula ϕ^* patrí do **S4** modálnej logiky. Z týchto skutočností vyplýva priamo možnosť použitia Kripkeho sémantického modelu pre intuicionistickú logiku, ktorý môžeme zaviesť analogickým spôsobom ako pre modálnu logiku (pozri podkapitolu 4): $W = (W, R, w_0, val)$, kde W je množina svetov, $R \subseteq W \times W$ je binárna reflexívna a tranzitívna relácia dostupnosti, je odlišený svet, ktorý je špecifický pre daný model, a $val: W \times P \rightarrow \{\mathbf{0}, \mathbf{1}\}$ je zobrazenie, ktoré vyhovuje nasledujúcej vlastnosti perzistentnosti pre každú výrokovú premennú $p \in P$:

- (1) ak $(v(w,p)=1)$, potom pre každé $w' \in \Gamma(w): (v(w',p)=1)$
 (2) ak $(v(w,p)=0)$, potom pre každé $w' \in \Gamma(w): (v(w',p)=0/1)$

Zobrazenie *val* je zovšeobecnené pre formuly ϕ a ψ takto:

(1)	ϕ je pravdivé pre svet w_0 v modeli M	vtt	ak ϕ je nepravdivé pre každý $w' \in \Gamma(w)$ v M
(2)	$\phi \wedge \psi$ je pravdivá pre svet w_0 v modeli M	vtt	ak ϕ je pravdivé pre w_0 v M a ψ je pravdivé pre w_0 v M
(3)	$\phi \vee \psi$ je pravdivá pre svet w_0 v modeli M	vtt	ak ϕ je pravdivé pre w_0 v M alebo ψ je pravdivé pre w_0 v M
(4)	$\phi \Rightarrow \psi$ je pravdivá pre svet w_0 v modeli M	vtt	ak ϕ je nepravdivé alebo ψ je pravdivé pre každý $w' \in \Gamma(w)$ v M

Riadky (1) a (4) v tejto tabuľke môžu byť jednoducho interpretované tak, že intuicionistické spojky \neg a \Rightarrow sú vyjadrené pomocou modálnej spojky ako $\Box\neg$ resp. $\Box\Rightarrow$. Dôkaz korektnosti a úplnosti Heytingovho systému intuicionistickej logiky vzhľadom k tomuto sémantickému modelu môže byť vykonaný úplne analogickým spôsobom ako bol vykonaný pre **S4** modálnu logiku, iba musí byť doplnený Gödelovým transformačným postupom uvedeným v prvej časti tejto kapitoly. Existuje aj priama alternatíva tohto dôkazu, ktorá bola navrhnutá Kripkem, kde Gödelov transformačný postup môže byť dokázaný ako jednoduchý dôsledok. Poznajme, že mnohé vlastnosti intuicionistickej logiky, ktoré sú v rámci Heytingovho systému dokázateľné s obtiažami, v rámci Kripkeho sémantického modelu tejto logiky sú dokázateľné ako jednoduché dôsledky vlastnosti korektnosti a úplnosti.⁶

9 Kripkeho model modálnej predikátovej logiky

Prvý, kto študoval kombináciu modálnej logiky s predikátovou logikou boli Ruth Barcanová (vydatá Marcusová) [24] a Rudolf Carnap [2], pričom Carnapov prístup nevyhovoval intuitívnym vlastnostiam, ktoré sa očakávali od kombinácie týchto dvoch logík a prístup Barcanovej bol čisto formálny na syntaktickej úrovni a nezaoberal sa sémantickou interpretáciou. V tejto kapitole naznačíme Kripkeho model [14] modálnej predikátovej logiky, ktorý vznikne kombináciou sémantických modelov predikátovej a modálnej logiky. Vzniknutý model súčasne obsahuje tak unárne spojky \Box , a \Diamond ako aj kvantifikátory \forall a \exists . Jazyk modálnej predikátovej logiky zostrojíme jednodu-

⁶ Najznámejšia z týchto vlastností je vlastnosť disjunkcie, že ak je teorém intuicionistickej logiky, potom buď ϕ alebo ψ je taktiež teorém. Dôkaz tejto vlastnosti na syntaktickej úrovni je pomerne zdĺhavý. Dôkaz tejto vlastnosti na sémantickej úrovni je triviálny.

cho tak, že jazyk predikátovej logiky rozšírime o obe modálne spojky. Sémantika takto vzniknutej modálnej predikátovej logiky [3] žiaľ už nie je taká jednoduchá ako je sémantika tak modálnej logiky (založenej na Kripkeho modeli dostupných svetov), ako aj predikátovej logiky [21]. Nové zovšeobecnenie spočíva v tom, že univerzum U (vzhľadom ku ktorému sú definované kvantifikátory) závisí od sveta $w \in W$ v ktorom je študovaná formula ϕ modálnej predikátovej logiky. Pre každý svet $w \in W$ vyberieme podmnožinu $U(w) \subseteq U$, ktorá špecifikuje kvantifikátory vzhľadom k danému svetu. Budeme uvažovať tieto tri alternatívy

$$(w, w') \in R \Rightarrow \begin{cases} U(w) = U(w') & (\text{konštantné univerzum}) \\ U(w) \subseteq U(w') & (\text{rastúce univerzum}) \\ U(w) \supseteq U(w') & (\text{klesajúce univerzum}) \end{cases}$$

Pre modálnu predikátovú logiku Kripkeho sémantický model (pozri definíciu Kripkeho modelu v podkapitole 4) budeme špecifikovať v zjednodušenom tvare

$$M = (W, R, v, w_0, \{U(w)\}_{w \in W})$$

kde $W = \{w, w', w'', \dots\}$ je **množina svetov** a $W \subseteq W \times W$ je binárna relácia **dostupnosti** nad množinou W . Výraz V špecifikuje **pravdivostnú interpretáciu** formúl modálnej predikátovej logiky, obsahuje tak pravdivostnú interpretáciu predikátovej logiky, ako aj modálnej logiky. Pretože prezentujeme zjednodušený sémantický model modálnej predikátovej logiky, nebudeme bližšie špecifikovať sémantický interpretačný postup obsiahnutý v operátore v . Výraz w_0 špecifikuje **vybraný svet** z W , vzhľadom ku ktorému je zostrojený model M . Konečne, $\{U(w)\}_{w \in W}$ je **postupnosť univerzálnych množín** špecifikovaných pre všetky možné svety z W . Predpokladáme, že tieto množiny vyhovujú jednej z troch vyššie uvedených relácií, t. j. sú konštantné, rastúce, alebo klesajúce.

Pre takto špecifikovaný sémantický model modálnej predikátovej logiky môžeme priamočiaro zovšeobecniť metódu sémantických tabiel. Pre ilustráciu uvidíme dva prípady rozvoja subformuly ak súčasne obsahuje kvantifikátor a modálnu spojku v kombináciách $\Box \forall$ a $\Box \exists$:

$$\begin{array}{ccc} & \downarrow & \downarrow \\ w_0 \models \Box(\forall x)P(x) & \rightarrow & w_0 \models \Box(\forall x)P(x) \\ & & \downarrow \\ & & w \models P(t) \end{array}$$

kde $w \in \Gamma(w_0)$ je ľubovoľný svet dostupný zo sveta w_0 a $t \in U(w)$ je ľubovoľný objekt z univerza $U(w)$ špecifikovaného pre svet w . Podobne dostaneme aj

$$\begin{array}{ccc} & \downarrow & \\ w_0 \models \diamond(\exists x)P(x) & \rightarrow & w_0 \models \diamond(\exists x)P(x) \\ & & \downarrow \\ & & w \models P(a) \end{array}$$

kde $w \in \Gamma(w_0)$ je vybraný svet dostupný zo sveta w_0 a $a \in U(w)$ je vybraný objekt z univerza $U(w)$ špecifikovaného pre svet w . Aj v tomto prípade platí známa veta sémantických tabiel, že formula φ je tautológiou v rámci Kripkeho modelu predikátovej logiky práve vtedy ak tablo $T(\neg\varphi)$ obsahuje len uzavreté vetvy, t. j. pre každú vetvu existuje kontradikcia pre atomické premenné. Ak tablo nie je uzavreté, potom daná otvorená vetva sa dá použiť na konštrukciu jednoduchého Kripkeho modelu pre ktorý daná formula je nepravdivá, čiže sa falzifikuje predpoklad jej tautologičnosti.

Takto špecifikovaná metóda sémantických tabiel môže byť použitá pre dôkaz toho, že tzv. Barcanovej formuly [27] sú tautológie pre rôzne prípady množín $U(w)$, špecifikovaných pre aktuálny svet $w \in \Gamma(w_0)$. Získané výsledky sú zhrnuté v nasledujúcej tabuľke:

#	Formula	konštantné univerzum	rastúce univerzum	klesajúce univerzum
(1)	$\diamond(\exists x)P(x) \Rightarrow (\exists x)\diamond P(x)$	áno	nie	áno
(2)	$(\forall x)\Box P(x) \Rightarrow \Box(\forall x)P(x)$	áno	nie	áno
(3)	$\Box(\forall x)P(x) \Rightarrow (\forall x)\Box P(x)$	áno	áno	nie
(4)	$(\exists x)\diamond P(x) \Rightarrow \diamond(\exists x)P(x)$	áno	áno	nie

Tak napríklad, Barcanovej formula $\diamond(\exists x)P(x) \Rightarrow (\exists x)\diamond P(x)$ je tautológia pre konštantné a klesajúce univerzum a nie je tautológia pre rastúce univerzum.

Kripkeho sémantický model pre predikátovú modálnu logiku môžeme chápať ako priamočiare zovšeobecnenie jeho modelu výrokovej modálnej logiky pre prípad, že logika obsahuje kvantifikátory, pričom univerzum U je špecifikované pre aktuálne svety z množiny \mathcal{W} . V po-

užitom ilustračnom príklade bol použitý podstatne zjednodušujúci predpoklad, že univerzá $U(w)$ sú buď konštantné, klesajúce alebo rastúce vzhľadom k binárnej relácii dostupnosti R . Použitý jednoduchý sémantický model predikátovej modálnej logiky spolu s uvedenou teóriou sémantických tabiel môže byť použitý k dôkazu úplnosti a korektnosti pre rôzne verzie tejto logiky [3]. Poznamenajme, že naznačený prístup pre konštrukciu sémantického modelu predikátovej modálnej logiky je použiteľný aj pre konštrukciu sémantického modelu predikátovej intuicionistickej logiky špecifikovanej v 8. podkapitole.

10 Záver

Predchádzajúce pokusy o zostrojenie analogickej idey ku Kripkeho sémantike umožňujú v súčasnosti vznik mnohých sporov o prioritu. Myšlienka použitia binárnej relácie k modelovaniu sa vyskytla u mnohých ľudí s rôznorodým odôvodnením. Je potrebné však zdôrazniť, že Kripke bol prvý, ktorý vytvoril relatívne jednoduchý sémantický model modálnej logiky, ktorý sa ukázal, napriek svojej jednoduchosti, ako veľmi plodný pre ďalší rozvoj neklasických logík. Jeho model možno charakterizovať ako množinovo-teoretický, zatiaľ čo iné modely využívali rôzne variácie na tému binárnych relácií medzi nemodálnymi fragmentmi predikátovej logiky, a taktiež nedávali vlastnosti binárnej relácie do vzájomného vzťahu s axiomatizáciou danej modálnej logiky. Kripke bol prvý, ktorý pomocou vlastností binárnej relácie umožnil v rámci sémantického prístupu modelovať rôzne „syntaktické“ modely modálnej logiky a tým dostal do vzájomného vzťahu rôzne axiomatické systémy modálnej logiky s ich jednoduchou charakteristikou pomocou vlastností binárnej relácie R .

PodĎakovanie: Tento príspevok vznikol za podpory grantovej agentúry VEGA v rámci grantových úloh 1/0804/08, 1/4053/07 a agentúry APVV v rámci grantovej úlohy 20-002504.

Literatúra

- [1] Beth, E. W.: Semantic Entailment and Formal Derivability. *Mededelingen der Koninklijke Nederlandse Akademie van Wetenschappen*. **18** (1955) 309-342.
- [2] Carnap, R.: Modalities and Quantification. *Journal of Symbolic Logic* **11** (1946) 33-64.
- [3] Garson, J. W.: *Modal Logic for Philosophers*. Cambridge University Press, New York 2006.
- [4] Goldblatt, R.: Mathematical Modal Logic: a View of its Evolution. *Journal of Applied Logic* **1** (2003) 309-392.
- [5] Gödel, K.: Eine Interpretation des intuitionistischen Aussagenkalküls. *Ergebnisse eines mathematischen Kolloquiums* **4** (1932) 39-40.
- [6] Gruska, J.: *Quantum Computing*. Mc-Graw Hill, London 1999.
- [7] Heyting, A.: *Intuitionism: An Introduction*. North Holland, Amsterdam 1956.

- [8] Hintikka, J.: The Modes of Modality. *Acta Philosophica Fennica* **16** (1963) 65-82.
- [9] Hintikka, J.: Is Alethic Modal Logic Possible? *Acta Philosophica Fennica* **35** (1982) 89-105.
- [10] Holmström-Hintikka, G., Lindström, S., Sliwinski, R. (eds.): *Collected Papers of Stig Kanger with Essays on his Life and Work*, Vol. 1. Kluwer, Dordrecht 2001.
- [11] Jónsson, B., Tarski, A.: Boolean Algebras with Operators. *American Journal of Mathematics* **73** (1951) 891-939.
- [12] Kanger, S.: *Provability in Logic*. Acta Universitatis Stockholmensis. Almqvist and Wiksell, Stockholm 1957.
- [13] Kripke, S.: A Completeness Theorem in Modal Logic. *Journal of Symbolic Logic* **24** (1959) 1-14.
- [14] Kripke, S.: The Undecidability of Monadic Modal Quantification Theory. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik* **8** (1962) 113-116.
- [15] Kripke, S.: Semantical Analysis of Modal Logic I. Normal Propositional Calculi. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik* **9** (1963) 67-96.
- [16] Kripke, S.: Semantical Considerations on Modal Logic. *Acta Philosophica Fennica* **16** (1963) 83-94.
- [17] Kripke, S.: Semantical Analysis of Intuitionistic Logic I. In Crossley, J. N., Dummett, M. A. E. (eds.): *Formal Systems and Recursive Functions*. North Holland, Amsterdam 1963, pp. 92-129.
- [18] Kripke, S.: Semantical Analysis of Modal Logic II. Non-Normal Modal Propositional Calculi. In Addison, J. W., Henkin, L., Tarski, A. (eds.): *The Theory of Model*. North Holland, Amsterdam 1965, pp. 206-220.
- [19] Kripke, S.: Identity and Necessity. In Munitz, M. K. (ed.): *Identity and Individuation*. University Press, New York 1971.
- [20] Kripke, S.: Naming and Necessity. In Davidson, D., Harman, G. (eds.): *Semantics of Natural Language*. Reidel, Dordrecht 1972, pp. 253-355.
- [21] Kvasnička, V., Pospíchal, J.: *Matematická logika*. Vydavatel'stvo STU, Bratislava 2006.
- [22] Lewis, C. I.: *A Survey of Symbolic Logic*. University of California Press, Berkeley 1918.
- [23] Lewis, D.: *On the Plurality of Worlds*. Basil Blackwell, Oxford (UK) 1986.
- [24] Marcus, R. B.: A Functional Calculus of First Order Based on Strict Implication. *Journal of Symbolic Logic* **11** (1946) 1-16.
- [25] McKinsey, J. C. C., Tarski, A.: Some Theorems About the Sentential Calculi of Lewis and Heyting. *Journal of Symbolic Logic* **13** (1948) 1-15.
- [26] Pavel, T.: *Fictional Worlds*. Harvard University Press, Cambridge, Mass. 1986.
- [27] Peregrin, J.: *Logika a logiky*. Academia, Praha 2004.

Modely jazyka ako dynamickej siete

Mária Markošová¹

Abstrakt. Pozornosť kognitívnych vied sa v súčasnosti upiera na modelovanie myslenia. Ak chceme porozumieť fungovaniu neurónovej siete mozgu, potrebujeme porozumieť fungovaniu sietí ako takých. Predložená práca sa zaoberá rastúcimi sieťami. Jednotlivé modely rastúcich sietí ukazujú, ako spôsob rastu siete a spôsob pripájania uzlov do siete ovplyvňuje jej celkovú štruktúru a vlastnosti. Modely sú aplikované na lexikón jazyka. Spomenuté sú aj niektoré ďalšie, najmä biologické, ale aj iné aplikácie teórie sietí.

1 Úvod

Modelovať myslenie človeka je veľmi zložitá úloha, ktorú si vytýčili kognitívne vedy. Na to, aby sme si vedeli predstaviť, ako myseľ pracuje, musia spojiť svoje sily vedci z rôznych oblastí. Myslenie je vlastnosť špecificky ľudská, aj keď zoológovia, ktorí sa zaoberajú primátmi, by proti tejto téze mohli mať isté námietky. Myslenie má svoje antropologické, sociologické i filozofické dôsledky. Myšlienky si ľudia vymieňajú pomocou jazyka. Fenomén jazyka nesie v sebe rôzne filozoficko - lingvistické problémy ako je, napríklad, vyjadrovacia schopnosť jazyka, schopnosť inej mysle porozumieť, problém presnosti jazyka i problém samotnej štruktúry jazyka. Myslenie však má aj svoj biologický základ – je vlastnosťou ľudského mozgu, a teda je aj výzvou pre biológov.

Mozog je vlastne zložitá sieť neurónov, ktorá vo forme elektrických impulzov prenáša vzruchy a informácie. Tým sa myslenie stáva aj informatickým a fyzikálnym problémom, pretože si môžeme klásť otázky o tom, aká je štruktúra tejto siete, aké má fyzikálne vlastnosti, ako dobre a ako rýchlo je schopná prenášať informáciu. Štúdium sietí má teda pre porozumenie činnosti ľudského mozgu veľký význam.

Ak chceme vedieť niečo o štruktúre sietí, je dobré najprv preskúmať jednoduché modely a až potom uvažovať, nakoľko modely zodpovedajú reálnym sieťam. V tejto práci rozoberiem niekoľko jednoduchších, analyticky riešiteľných modelov rastúcich sietí [1-4]. Najprv sa budem zaoberať klasickými modelmi siete s náhodným a preferenčným pripájaním uzlov a ukážem, ako spôsob pripájania uzlov ovplyvňuje výsledné vlastnosti siete [1]. V ďalšej kapitole sa pozriem na štruktúru lexikónu jazyka [2,5,6]. Zavedením dodatočných procesov do dynamickej siete ukážem, že je možné vytvoriť model zodpovedajúci reálnym jazykovým dátam [6]. V poslednej kapitole pouká-

¹ Katedra aplikovanej informatiky, Fakulta matematiky, fyziky a informatiky UK, Bratislava, e-mail: markosova@ii.fmph.uniba.sk

žem na možné aplikácie modelov aj z hľadiska modelovania neurónovej siete mozgu.

1.1 Rastúce siete

Na začiatok skúsme spomenúť zopár príkladov z reality. Predstavme si, napríklad, telefónnu sieť. Telefóny boli najprv vzájomne poprepájané po dvojiciach. Ak chcel jeden priateľ volať inému, museli si obaja kúpiť prístroje a tie medzi sebou spojiť. Neskôr sa objavili telefónne ústredne, ktoré boli vzájomne prepojené a jednotlivé prístroje sa napájali na ústredňu. Tu už môžeme hovoriť o hierarchickej sieti, ktorá rástla jednak pridávaním nových ústrední, jednak pripájaním nových uzlov, telefónnych prístrojov, k ústredniam.

Ďalším príkladom môže byť sieť sociálnych alebo profesionálnych kontaktov [7,8]. Každý z nás má svoj lokálny klaster priateľov a známych. Cez nového priateľa zasa môžeme teoreticky získať prístup k celému klastru jeho známostí. Ak sa niekoľko z našich známych odsťahuje do vzdialených končín sveta a nepreruší s nami kontakt, vytvára tým výhodné „krátke spojenie“, cez ktoré sa môžeme dostať do styku s ľuďmi, ku ktorým by sme sa inak nikdy nedostali.

Všetci sedávame pri počítačoch, ktoré sú obvykle pripojené na internetovú sieť. Premýšľame pomocou neurónovej siete nášho mozgu, používame mobilnú sieť, a v našom tele prebiehajú procesy na rôznych interakčných biologických sieťach .

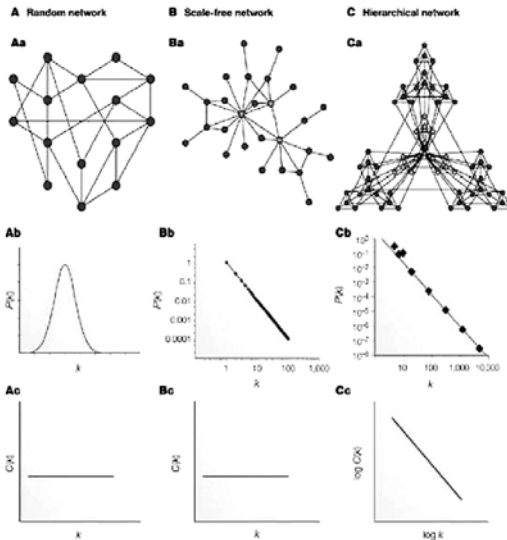
Väčšina týchto sietí sa mení s časom. Ak do nich uzly len pribúdajú, alebo ak pribúdanie uzlov výrazne prevyšuje zánik uzlov, hovoríme o rastúcich sieťach. Uzly sa do siete môžu pripájať rôznym spôsobom. Asi intuitívne tušíme, že štruktúra siete bude iná, ak je v nej istá hierarchia, tak ako napr. v telefónnej sieti, alebo ak sa nový uzol pripojí ku starému uzlu úplne náhodne.

Predstavme si najjednoduchšiu sieť, v ktorej hrana medzi dvoma uzlami buď existuje, alebo nie (binárne siete) a ak existuje, nemá význačný smer (nesmerované siete). Pod stupňom k_i uzla i rozumieme počet hrán, ktorými je uzol spojený so svojimi susedmi. Najjednoduchšou mierou, ktorá niečo hovorí o štruktúre siete, je priemerný stupeň uzla \bar{k} . Táto miera je ale dobrá len vtedy, ak v sieti existuje typický uzol s typickým stupňom, ako je tomu napríklad v **náhodných sieťach** (obr. 1) Erdosa a Renyiho [4]. Distribúcia stupňa uzlov v takejto sieti má peak. Znamená to, že z veľkej väčšiny uzlov vychádza približne ten istý počet hrán, ktorý je blízky \bar{k} .

Je všeobecne známe [3,4,8], že väčšina reálnych sietí, vznikajúcich samoorganizáciou, má tzv. **bezškálovú štruktúru** (obr. 1), vyjadrenú mocninným zákonom pre distribúcie stupňa uzlov. Pravdepodobnosť

toho, že uzol má práve stupeň k , je v bezškálových sieťach daná ako $P(k) \propto k^{-\gamma}$. To znamená, že v sieti neexistuje žiaden význačný uzol, ale že štruktúra siete je sebedobná na rôznych škálach. Ako sa takáto bezškálová štruktúra v sieti vytvorí? Procesy, ktorými vzniká, musia byť veľmi jednoduché. Usudzujeme tak preto, že výskyt bezškálových sietí je tak častý, že musia rásť prirodzeným samoorganizujúcim sa spôsobom alebo procesom.

Aký je to proces? Ako príklad nám môže poslúžiť komunikačná sieť, napríklad sieť železníc v krajine. Keď plánovači plánujú novú železničnú trať, vždy sa ju snažia pripojiť k nejakému komunikačnému uzlu, kde sa zbíha veľké množstvo iných tratí. To totiž umožňuje, aby sa cestujúci ľahšie dostali tam kam chcú bez častého presadania. Podobná situácia je aj v sieťach sociálnych vzťahov. Človek, ktorý už má množstvo kontaktov, ľahšie nadviaže ďalšie. Môžeme sa preto opýtať: je preferovanie dobre prepojených uzlov siete novými uzlami tým faktorom, ktorý vedie na bezškálovú štruktúru siete? Súvisí globálna štruktúra siete s lokálnymi procesmi pripájania uzlov? Ako sa štruktúra siete mení, keď vznikajú, alebo zanikajú nové hrany medzi starými uzlami? V nasledujúcich častiach naznačím odpovede na niektoré z týchto otázok.



Nature Reviews | Genetics

Obrázok 1. Rôzne typy sietí. Aa – náhodná sieť a k nej prislúchajúce distribúcie stupňa uzlov Ab a priemerného klasterizačného koeficientu pre uzol stupňa k , Ac, B, C – to isté pre bezškálovú a pre hierarchickú sieť. Podľa [16].

1.2 Sieť s náhodným pripájaním uzlov

Predstavme si takýto proces:

- Na začiatku máme malú sieť, o malom počte uzlov, ktoré sú prepájané hranami, napríklad spôsobom každý uzol s každým. Na tom vlastne príliš nezáleží, pretože ak necháme sieť rásť dostatočne dlho, na jej vlastnosti to nebude mať vplyv.
- Nech každú časovú jednotku pribudne do siete jeden uzol, ktorý sa pripojí k starým uzlom m hranami. Nech pravdepodobnosť toho, že sa nový uzol pripojí k starému bude úmerná $\frac{1}{t}$ a nech t je čas meraný v nejakých časových jednotkách. Všimnime si, že t je vlastne počet uzlov, ktoré sa k sieti už pripojili a pravdepodobnosť, že starý uzol zachytí jednu z m nových hrán je rovnaká pre všetky staré uzly.

Nech $k(s, t)$ je priemerný stupeň uzla, ktorý prišiel do siete v čase s (a s bude preto aj označením tohto uzla) a pozorujeme ho v čase t . Potom pre náš proces platí

$$\frac{\partial k(s, t)}{\partial t} = \frac{m}{t} \quad (1)$$

Rovnica vlastne hovorí o tom, ako rýchlo sa s časom mení stupeň uzla s . Riešením tejto rovnice je

$$k(s, t) = m + m \ln\left(\frac{t}{s}\right) \quad (2)$$

Vidíme, že stupeň uzla s rastie s časom len pomaly.

Rovnica (2) je vyjadrením lokálnych pomerov v sieti. O jej globálnej štruktúre lepšie vypovedajú iné miery, napríklad distribúcia stupňa uzlov a separácia uzlov. Pod separáciou uzlov rozumieme priemernú najkratšiu vzdialenosť medzi dvoma náhodne vybranými uzlami v sieti. Distribúcia stupňa uzlov je funkcia, ktorá meria počet uzlov, ktoré majú stupeň k . Keď ju prenormujeme, dostaneme pravdepodobnosť nájdenia uzla so stupňom k pri náhodnom výbere [3]:

$$P(s, t) = \frac{1}{t} \int_0^t ds \delta(k - k(s, t)) = - \left[t \frac{\partial k(s, t)}{\partial t} \right]_{s=s(k, t)}^{-1} \quad (3)$$

Z rovnice (3) vieme, že ak poznáme ako sa bude s časom vyvíjať $k(s, t)$, potom dokážeme analyticky vypočítať distribúciu. Keď použijeme rovnice (2) a (3), dostaneme distribučnú funkciu siete s náhodným pripájaním uzlov [3]

$$P(k, t \rightarrow \infty) \propto \exp(-\alpha k). \quad (4)$$

Vidíme, že v sieti prakticky neexistujú uzly s veľkým stupňom, sieť pozostáva z veľkého množstva uzlov, ktoré sú do nej zapojené malým počtom hrán. Ale zároveň neexistuje ani typický uzol, s typickým stupňom.

1.3 Barabási – Albert model s preferenčným pripájaním uzlov

Roku 1999 vyšiel v Science článok Emergence of Scaling in Random Network [1], v ktorom bol navrhnutý model rastúcej siete s preferenčným pripájaním nových uzlov. Cieľom autorov bolo nájsť taký proces, ktorý by viedol na často pozorovanú bezškálovú štruktúru siete. Rozhodli sa preto preskúmať, či by týmto faktorom nemohol byť vyššie zmienený a v reálnych sieťach často pozorovaný princíp „ten, čo už má veľa susedov, bude ich mať ešte viac“, teda preferenčné pripájanie nových uzlov.

Pod preferenčným pripájaním uzlov rozumieme takýto proces:

- Na začiatku máme opäť malú sieť, o malom počte uzlov, ktoré sú, poprepájané hranami, napríklad spôsobom každý uzol s každým.
- Každú časovú jednotku pribudne do siete uzol a m hranami sa pripojí k starším uzlom. Pravdepodobnosť toho, že starší uzol zachytí novú hranu, je však v tomto prípade úmerná jeho stupňu. Čiže ide práve o situáciu, keď “víťaz berie viac”. V reálnych sieťach je to dosť častý jav. Napríklad v sieti profesionálnych kontaktov sú ľudia, ktorí sú veľmi známi a majú množstvo kontaktov. Keďže ich každý pozná, to, že sa na nich niekto nakontaktuje, je omnoho viac pravdepodobné, ako že sa nakontaktuje na málo známeho profesionála.

Ak sieť rastie dostatočne dlho, možno tento proces popísať, modelovať, diferenciálnou rovnicou

$$\frac{\partial k(s,t)}{\partial t} = m \frac{k(s,t)}{\int_0^t du k(s,t)} \quad (5)$$

Rovnica popisuje rýchlosť zmeny stupňa uzla, ktorý prišiel do siete v čase s , s časom. Pravá strana predstavuje pravdepodobnosť pripojenia nového uzla k uzlu s . Riešenie tejto rovnice je jednoduché, ak si uvedomíme, že integrál v menovateli predstavuje vlastne sumu všetkých stupňov uzlov, ktorá, vzhľadom na to, že počet uzlov je t , bude $2mt$. V systéme je totiž mt hrán, a z nich každá spája dva uzly. Riešenie (5) vyzerá takto:

$$k(s,t) \propto \left(\frac{t}{s}\right)^\beta \quad (6)$$

kde $\beta = 1/2$, a použitím (3) dostávame distribučnú funkciu stupňa uzlov

$$P(k, t \rightarrow \infty) \propto k^{-\gamma}, \quad (7)$$

kde $\gamma = 3$. Ak chceme zistiť len to, aký bude exponent, môžeme si ušetriť počítanie podľa rovnice (3), pretože platí [3, 4]

$$\gamma = 1 + \frac{1}{\beta} \quad (8)$$

Necháme teraz matematiku matematikou a skúsme sa zamyslieť, čo vlastne tieto výsledky znamenajú. V prvom rade to, že globálne vlastnosti siete, vyjadrené distribučnými funkciami, sú radikálne ovplyvnené dynamikou siete. Lokálne procesy pripájania uzlov tak určujú globálnu štruktúru siete. Na druhej strane, v konkrétnej sieti možno zmerať počet uzlov s daným stupňom, teda distribúciu. Z jej tvaru môžeme odhadnúť, ako sieť rástla, či sa do nej uzly pripájali prevažne preferenčným, alebo prevažne náhodným spôsobom (obr. 1).

Samozrejme, existujú aj iné modely. V podstate sú to variácie na dva základné typy, ktoré som práve popisala. Čitateľa, ktorý by sa o nich chcel dozvedieť viac, odkazujem na ďalšiu literatúru [3, 4].

1.4 Sieť jazyka

Teória je dobrá vtedy, keď poskytuje možnosť lepšie porozumieť realite. Dobrá teória rastúcich sietí by preto mala vedieť vysvetliť, aké lokálne procesy vedú k pozorovanej globálnej štruktúre. Aplikácií teórie rastúcich sietí je mnoho. Skúma sa štruktúra internetu [10], www siete [8], vlastnosti biologických sietí [11], sociálnych sietí [7-9], citačných sietí [8], komunikačných sietí [8], ale napr. aj jazyková sieť [2,6,7,12].

Pohľad na lexikón jazyka ako na sieť vzťahov medzi slovami je iste zaujímavý. Možno nám pomôže odhaliť niečo zo štruktúry lexikónu. Na druhej strane sieťový pohľad na jazyk by mohol pomôcť objasniť niektoré poruchy reči, keď napríklad reč stráca plynulosť (rozpad jazykovej siete), alebo nahrádzanie pojmov, ktoré človeku vypadli.

Jazykovú sieť možno vytvoriť takto: považujeme slová za uzly siete a vzťahy medzi slovami za hrany. Pod vzťahom medzi slovami môžeme rozumieť buď pojmovú príbuznosť [12], alebo polohu slova vo vete, teda susedstvo slov [5]. Modely, ktoré v tejto kapitole predstavím, sa zakladajú na druhom type vzťahu. Slovo pokladáme spojené hranami so všetkými tými slovami, ktoré sú v skúmanom texte jeho bezprostrednými susedmi vo vete. Takto vznikne pozičná jazyková sieť [2,5,6].

Autori [5] skúmali takúto sieť, vytvorenú na báze textov anglického národného korpusu. Ich sieť mala okolo 450 000 uzlov, bola teda dosť veľká. Po zmeraní distribúcie stupňa uzlov sa zistilo, že sieť rástla pravdepodobne tak, že jednotlivé slová – uzly sa k nej pripájali preferenčným spôsobom. Na rozdiel od (7) však v sieti možno pozorovať dva režimy. Pre malé hodnoty stupňa uzlov k je exponent $\gamma = 1.5$, zatiaľ čo pre veľké hodnoty k je γ blízke k teoreticky predpovedanej hodnote $\gamma = 3$.

Čo sa z toho dá usúdiť? Možno to, že okrem preferenčného pripájania uzlov existujú zrejme aj iné procesy, ktoré ovplyvňujú globálnu štruktúru siete. Nasledujúce modely jazykových sietí sa ich snažia zohľadniť.

2 Modely jazykových sietí

Rozdiel nameraných dát Cancha a Solého [5] od Barabásiho – Albert modelu [1] (5,6) sa pokúsili vysvetliť Dorogovtsev s Mendesom [2]. Navrhli ďalší proces. Je to tvorba nových hrán medzi starými uzlami, pričom uzly, medzi ktorými sa vytvorí hrana sa vyberajú preferenčným spôsobom. V reči opisujúcej lexikón jazyka to znamená, že niektoré slová sa s časom používajú v širšom kontexte, a dostávajú sa aj do susedstva iných slov ako predtým. Model Dorogovtseva a Mendesa je teda obohatením Barabási – Albert modelu, v ktorom zároveň s preferenčným pripojením nového uzla k m starým uzlom sa vytvorí aj ct nových hrán medzi starými uzlami, pričom konštanta $c < 1$. Každá hrana má dva konce, a teda sa vytvorí $2ct$ nových koncov hrán. Dorogovtsevov - Mendesov model vyzerá takto:

$$\frac{\partial k(s,t)}{\partial t} = (m+2ct) \frac{k(s,t)}{\int_0^t du k(u,t)} \quad (9)$$

K uzlu s sa teda s istou pravdepodobnosťou preferenčne pripojí buď hrana od novoprišlého uzla, alebo sa vytvorí nová hrana. Riešením rovnice (9) je [2]

$$k(s,t) \propto \left(\frac{t}{s}\right)^{1/2} \left(\frac{2m+ct}{2m+cs}\right)^{3/2} \quad (10)$$

a použitím (3), dostaneme distribučnú funkciu, v ktorej pre malé hodnoty stupňa uzlov je exponent $\gamma = 1.5$ a pre veľké hodnoty $\gamma = 3$. Výsledky teda kvalitatívne dobre modelujú namerané dáta. Kvalitatívne, ale nie kvantitatívne. Cancho a Solé [5] totiž v strnšej časti distribučnej krivky namerali $\gamma = 2.7$, a nie $\gamma = 3.0$.

Naše vlastné merania pozičnej jazykovej siete potvrdzujú fakt, že skutočné exponenty sú o niečo nižšie ako tie, ktoré predpovedá model Dorogovtseva a Mendesa [2] (obr. 2). Zaoberali sme sa starým textom, Bibliou, a skúmali sme rôzne anglické preklady Biblie, od tých najnovších po najstaršie. Jednotlivé verzie prekladov sa dajú nájsť na internete [17]. Z každého prekladu sme vytvorili pozičnú jazykovú sieť a zmerali sme distribúciu stupňa uzlov. Dostali sme podobné výsledky ako Cancho a Solé [5], len s tým rozdielom, že naše merania ukazujú $\gamma = 1.32$ v miernejšej, resp. $\gamma = 2.21$ v strmšej časti distribučnej krivky [13].

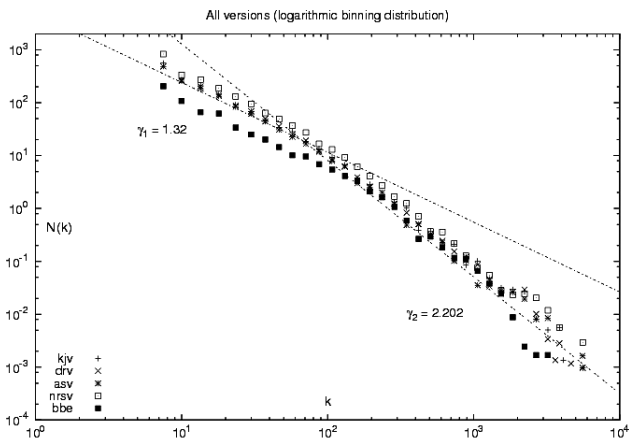
Je preto namieste usudzovať, že model (9) stále nie je úplný. Štruktúru siete pravdepodobne ovplyvňujú ďalšie procesy, ktoré v modeli zatiaľ nie sú zahrnuté. Aké by mohli byť? Skúsme sa trochu zamyslieť.

Kontext niektorých slov sa s časom nielen obohacuje, ale aj ochudobňuje. Význam slov sa dokonca môže značne zmeniť, a preto sa vyskytujú vo vete s inými susedmi. V jazyku sietí to znamená, že niektoré hrany môžu byť od slova odpojené a prelinkované na iné slovo. Navrhla som teda takýto proces [6]:

- Každú časovú jednotku do siete prichádza uzol a preferenčne sa pripojí m hranami k starým uzlom.
- Zároveň sa vytvorí ct nových hrán medzi starými uzlami a preferenčne sa k nim pripoja.
- A v tom istom čase sa náhodne vyberie m_r starých uzlov, od každého z nich sa odpojí jedna hrana a prelinkuje sa k inému uzlu preferenčným spôsobom.

Môj model jazykovej siete vyzerá teda takto [6]:

$$\frac{\partial k(s,t)}{\partial t} = (m+2ct+m_r) \frac{k(s,t)}{t} - \frac{m_r}{t} \int_0^t du k(u,t) \quad (11)$$



Obrázok 2. Distribúcia stupňa uzlov pre pozičnú slovnú sieť, ktorá bola skonštruovaná na báze anglických prekladov Biblie. Niektoré preklady sú staršie (Douay Rheims verzia, djv, rok vydania 1582; King James verzia, kjv, 1611), iné moderné (American Standard verzia, asv, 1901; Basic English verzia, bbe, 1941; New Revised Standard verzia, nrsv, 1989. bbe je špeciálny prípad, lebo text bol umelo zjednodušaný). Siete majú tento počet uzlov: drv – 11423, kjv – 11624, asv – 10105, nrsv – 14985, bbe – 4961.

a obohacuje predošlý model (9) o proces prelinkovanie hrán. V tomto modeli sa stupeň uzla s môže zvýšiť tak, že sa k nemu pripojí niektorá z m hrán novoprišlého uzla, alebo sa vytvorí nová hrana medzi ním a niektorým zo starších uzlov, alebo sa k nemu pripojí koniec hrany, ktorý sa odpojil od iného uzla. Stupeň uzla s sa môže aj znížiť, ak sa od neho odpojí niektorá hrana a prelinkuje sa inam.

Pretože integrál sumuje stupne všetkých uzlov, proces prelinkovania hrany ho neovplyvňuje a jeho hodnota je taká istá ako pre model Dorogovtseva a Mendesa. Riešenie (11) vyzerá takto [6]:

$$k(s, t) \propto \left(\frac{t}{s}\right)^A \left(\frac{2m+ct}{2m+cs}\right)^{2-A} \quad (12)$$

kde $A = \frac{m+m_r}{2m}$. To znamená, že exponent $\gamma = 2 + \frac{m-m_r}{m+m_r}$ (8) pre strmšiu časť distribúcie a $\gamma = 1.5$ pre menej strmú časť [6]. Ak zvolíme $10 \geq m \geq m_r > 0$, potom pre exponent v strmšej časti platí $3 > \gamma \geq 2$. Vzhľadom na to, že nepredpokladáme že m v jazykovej sieti bude väčšie ako 10, už malý počet prelinkovaných hrán spôsobí systematický pokles exponentu v strmšej časti distribučnej krivky.

Nový model teda vysvetlil ďalší rozdiel medzi nameranými dátami a Barabási – Albert modelom [1]. Nevysvetlil však všetky roz-

diely. Je možné, že sieť treba obohatiť o nové procesy, poprípade parametre, ktoré budú mať vplyv na jej celkovú štruktúru. O aké parametre by mohlo ísť, to je predmetom ďalšieho výskumu.

Čo teda vieme doteraz? Vieme, že pozičná jazyková sieť pravdepodobne rástla tak, že jednotlivé slová – uzly sa do nej pripájali preferenčne. Slovo, ktoré sa vyskytuje v množstve kontextov, má veľkú šancu dostať sa do kontextu, susedstva vo vete aj s novým slovom. Samozrejme, aj slová, ktoré existujú v slovníku dlho, sa môžu dostávať do iných kontextov – toto reprezentuje proces tvorby nových hrán, medzi starými slovami. Niektoré slová menia význam, staré kontexty sa strácajú a slová sa začnú používať tak, ako sa predtým nepoužívali. V modeli jazykovej siete je tento fakt reprezentovaný prepájaním jednotlivých koncov hrán od pôvodných slov na iné.

3 Ďalšie aplikácie teórie sietí

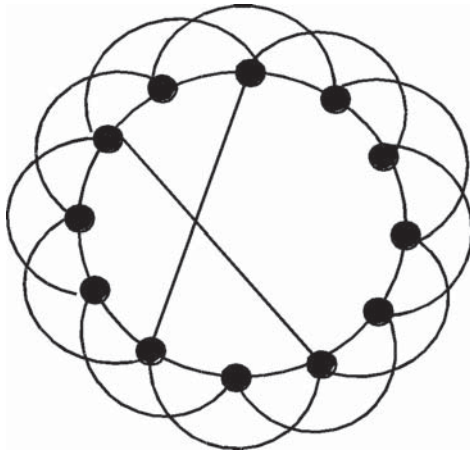
Procesy v sieťach, ktoré sme práve popísali, nie sú viazané len na jazykovú aplikáciu. Podobné procesy sa môžu vyskytnúť aj v iných sieťach. Keď zmeriame rozdelenie stupňa uzlov v nejakej reálnej sieti, môžeme z jeho tvaru a na základe znalostí získaných z modelov odhadnúť, aká bola história siete.

Almaas a Barabasi študovali štruktúru niektorých biologických sietí [11]. Dokázali, že biologické siete, ako napríklad metabolická sieť, proteínová interakčná sieť *S. cerevisiae* a rôzne siete potravinovej závislosti sú charakterizované mocninnými zákonmi typu (7). Títo autori zistili, že mnohé biologické siete sú jednak bezškálové (7), jednak **hierarchické** a niektoré možno nazvať aj **sieťami malého sveta** [8]. Z prvej kapitoly už vieme, čo je bezškálová sieť. Pojem „hierarchická sieť“ a „sieť malého sveta“ je však potrebné objasniť.

Jednou z mier, pomocou ktorej charakterizujeme siete na lokálnej úrovni, je aj klasterizačný koeficient C_i uzla i . Je to pomer počtu aktuálne existujúcich hrán medzi susedmi uzla i , k možnému počtu hrán medzi susedmi. Ak $C_i=1$, potom uzol i patrí k dokonale prepojenému klastru, ak $C_i=0$, uzol i je mostom medzi dvoma klastrami uzlov. Klasterizačný koeficient siete C predstavuje priemernú hodnotu C_i , spriemernú cez všetky uzly siete. Tak ako priemerný stupeň uzla, ani klasterizačný koeficient siete nemá sám osebe veľkú výpovednú hodnotu. Lepšou mierou je $C(k)$. $C(k)$ je priemerný klasterizačný koeficient uzlov, ktoré majú stupeň k a je mierou modularity siete. Ak platí, že

$$C(k) \propto k^{-\delta} \quad (13)$$

potom v sieti existuje **hierarchia** (obr. 1). V prípade, že v sieti platí zároveň (7) a (13), potom hovoríme o hierarchickej bezškálovej sieti. K takým patria všetky známe metabolické siete, napr. metabolická sieť *E. coli* [11].



Obrázok 3. Newmanova – Wattsova sieť malého sveta. Pravidelná štruktúra je narušená niekoľkými náhodnými ďalekodosahovými prepojeniami.

Pod pojmom „sieť malého sveta“ (obr. 3) rozumieme takú sieť, ktorá v sebe integruje vlastnosti usporiadanej siete s niektorými výhodnými vlastnosťami náhodnej siete. V akom zmysle? Klasterizačný koeficient siete C , ktorý je mierou lokálnej štruktúry, je v sieťach malého sveta pomerne veľký ($0.5 \leq C < 1$). Na druhej strane separácia uzlov je malá, porovnateľná so separáciou uzlov v náhodných grafoch. Je tomu tak vďaka niekoľkým náhodným krátkym spojeniam, ktoré prepojujú rôzne časti siete. Malá separácia uzlov je často výhodná vlastnosť, lebo umožňuje dobrú komunikáciu medzi uzlami siete a rýchle šírenie signálov.

Otázka, či jazyková sieť je bezškálovou a zároveň aj hierarchickou sieťou, je v štádiu výskumu. Je dosť pravdepodobné, že takou bude. Hierarchia totiž nie je v reálnych sieťach vzácnym úkazom. Ravasz a kol. [14] ju popisali ako mnohonásobnú iteratívnu multiplikáciu istej štruktúry. Vieme tiež, že jazyková sieť je sieťou malého sveta [5]. Určite by bolo veľmi zaujímavé preskúmať, aká je z hľadiska kombinácie hierarchie a bezškálovosti neurónová sieť ľudského mozgu. Po neurónovej sieti mozgu sa signály šíria rýchlo, preto separácia uzlov musí byť malá. O neurónovej sieti mozgu vieme aj to, že má modulárnu štruktúru. Predpokladám preto, vzhľadom na funkčné

vlastnosti mozgu, že to bude bezškálová sieť s vlastnosťami siete malého sveta a s istou hierarchickou modularitou.

Keďže biologické siete sú často hierarchickými a zároveň bezškálovými sieťami, majú štruktúru, ktorá je odolná voči náhodným poruchám, ale veľmi citlivá na cieľené ataky [4]. Takéto siete pozostávajú len z malého počtu uzlov s vysokým stupňom. Cieľený atak, teda cieľené vymazanie niektorého z takýchto uzlov, s veľkou pravdepodobnosťou spôsobí rozpad siete. Avšak pri náhodnom výbere je len málo pravdepodobné, že vyberieme uzol s veľkým stupňom a sieť sa rozpadne. Štruktúra biologických sietí preto dobre odoláva náhodným vplyvom, ale biologické siete, a s nimi aj biologická rovnováha procesov, ktoré sa dejú napríklad na interakčných biologických sieťach, je veľmi citlivá na devastáciu svojich kľúčových prvkov. Ak človek nechce drasticky narušiť túto citlivú rovnováhu, tak by to pri svojej činnosti mal brať do úvahy. Celá vec sa ešte viac komplikuje tým, že biologické siete nemožno považovať ani za binárne, ani za nesmerované.

Ďalšou zaujímavou aplikáciou teórie sietí sú siete sociálnych kontaktov. Je to vlastne jedna z prvých aplikácií; sieťovou štruktúrou sociálnych vzťahov sa zaoberal už Millgram roku 1969 [7]. Problémom však je ako objektívne a bezrozporne definovať sociálny kontakt. Nie je to jednoduché. V niektorých sieťach profesionálnych kontaktov sa ešte dá nájsť objektívna definícia. Môžeme napr. povedať, že ak herci hrajú spolu v jednom filme, sú v profesionálnom kontakte. Ale v bežných medziľudských vzťahoch platí, že ak niekoho poznám z videňa, je to iná úroveň vzťahu, ako keď sa s niekým kamarátim, alebo dokonca priateľim. A hranice medzi kamarátstvom a priateľstvom sú tiež neostre a subjektívne. Preto sa Liljevoss a spol. rozhodli študovať iný typ sociálnych kontaktov, sexuálne kontakty [9], pretože netrpia vyššie spomenutou nejednoznačnosťou. V sieti sexuálnych kontaktov človek je uzol a hrana medzi dvoma ľuďmi vznikne vtedy, ak mali spolu pohlavný styk.

Autori [9] ukázali, že distribúcia počtu sexuálnych partnerov má mocninný charakter (7). Zdá sa preto, že sieť sexuálnych kontaktov je bezškálovou sieťou (7). Pretože hrany v sieti sexuálnych kontaktov nie sú stabilné, ale s časom vznikajú i zanikajú, autori skúmali sieť kontaktov jednak za dobu jedného roku, jednak za celý doterajší život, a to na vzorke 4781 Švédov vo veku od 18 do 74 rokov. Pre krátkodobé kontakty predpokladali, že v krátkom časovom horizonte jedného roku k veľkým zmenám nedôjde (čo, samozrejme, nie je vždy pravda) a sieť je teda relatívne stabilná. Skúmali tiež ako sa distribúcia stupňa uzlov (7) mení, ak vezmeme do úvahy počet sexuálnych kontaktov za celú

dobu doterajšieho života. Autori [9] zistili, že ku kvalitatívnym zmenám nedochádza. Čo sa mení, je škálovací exponent γ (7). V prípade krátkodobej siete je $\gamma = 1.51$ (ženy) a $\gamma = 1.31$ (muži). Dlhodobé merania ukazujú $\gamma = 1.1$, resp. $\gamma = 0.6$ pre ženy, resp. mužov.

Pretože táto publikácia sa týka modelov mysle, spomeniem ešte jeden príklad. V práci Critical brain networks [15] autor ukazuje, že funkcionálne siete v mozgu majú tiež vlastnosť bezškálovosti (7) s exponentom $\gamma \approx 2.0$. Pod funkcionálnou sieťou sa rozumie sieť definovaná nie skutočnými prepojeniami rôznych oblastí mozgu, ale koreláciami aktivít medzi malými, dopredu definovanými časťami mozgu – tzv. voxelmi. Mierou korelácie je korelačný koeficient. Ak korelačný koeficient dosahuje istú, vopred určenú prahovú hodnotu, potom sa dva voxely považujú za prepojené linkou. Takto vytvorená sieť má ešte okrem bezškálovej štruktúry aj štruktúru siete malého sveta [8], kde malá separácia uzlov je doprevádzaná pomerne veľkou klasterizáciou. Ako už vieme, klasterizáciu meriame pomocou sieťového klasterizačného koeficientu a separácia je daná priemernou najkratšou vzdialenosťou medzi uzlami.

Čo z toho vlastne vyplýva? Zdá sa, že mozog vytvára funkčné siete, ktoré sú lokálne pomerne dobre prepojené, majú dobre definovanú lokálnu štruktúru, ale napriek tomu jednotlivé uzly nemajú k sebe ďaleko.

Modelovanie myslenia ako procesov na sieťach je ešte len v začiatkoch. Je to neprebádané pole tak pre biológov, ako aj pre fyzikov a informatikov. Domnievam sa, že dobrá znalosť procesov na sieťach by do veľkej miery mohla pomôcť pochopiť záhady a limity ľudskej mysle.

4 Diskusia

V tomto článku som na príklade modelov rastúcich sietí ukázala, ako súvisí štruktúra siete s procesmi, pomocou ktorých sieť rastie. Modely som aplikovala na pozičnú jazykovú sieť. Modely však majú všeobecnú platnosť, modelujú ktorékoľvek reálne siete s podobnou dynamikou.

Môj model jazykovej siete [6] síce dobre vysvetľuje niektoré disproporcie medzi nameranými dátami a predošlými modelmi [1,2], nevysvetľuje však všetko, a preto ho nemožno považovať za kompletný. Príčina zostávajúcich odlišností je vecou ďalšieho výskumu. Takisto nevieme, nakoľko je pozičná jazyková sieť organizovaná hierarchicky.

V budúcnosti by bolo veľmi zaujímavé sledovať, ako sa buduje pozičná slovná sieť v mysli človeka pri učení sa cudzieho jazyka. Ras-

tie ako celok pripájaním nových uzlov – slov k centrálnemu klastru, alebo vzniká prepojením jednotlivých, na začiatku učenia izolovaných, klastrov? Ako súvisí schopnosť človeka plynule hovoriť cudzím jazykom s momentom prepojenia týchto klastrov? To všetko sú otázky, ktoré ešte len čakajú na vyriešenie.

Jazyková sieť je však len jednou z množstva možných aplikácií teórie sietí. Teória sietí môže pomôcť porozumieť takým problémom kognitívnej vedy ako je napríklad myslenie. Mozog je vo svojej podstate komplikovaná sieť neurónov. Možno keď hlbšie porozumieme tomu, na akom type sieťovej štruktúry myslenie prebieha, porozumieme lepšie aj samotnému procesu myslenia.

Literatúra

- [1] Barabási, A. L., Albert R. : Emergence of scaling in random network, *Science*, **286**, (1999) 509 - 512
- [2] Dorogovtsev, S. N., Mendes, J. F. F. : Language as an evolving web, *Proc. Royal Soc. London B*, **268**, (2001) 2603 – 2606
- [3] Dorogovtsev, S. N., Mendes, J. F. F. : Evolution of networks, *Adv. Phys.* **51**, (2002) 1709 – 1770
- [4] Albert R., Barabási, A. L. : Statistical mechanics of complex networks, *cond-mat 0106096* (2001) 1 – 48²
- [5] Cancho, R. F., Solé, R. : Language as an evolving word web, *Proc. Royal Soc. London B*, **268**, (2001) 2261 – 2265
- [6] Markošová, M. : Network model of human language, *Physica A* **387**, (2008) 661 – 666
- [7] Milgram, S. : The small world problem, *Psychology Today* **2**, (1969) 60 – 67
- [8] Watts, D. : Small worlds, Princeton university press, Princeton, USA, ISBN 0-691-11704-7, (2004)
- [9] Liljeros, F. L., Edling, Ch. R., Amaral, L. A. N., Stanley E. : The web of human sexual contacts, *Nature* **411**, (2001) 907 – 908
- [10] Yook, S., Jeong, H., Barabási, A. L. : Modeling the Internet's large scale topology, *cond. mat.* 0107417 (2001) 1 – 15
- [11] Almaas, E., Barabási, A. L. : Power laws in biological networks, in Power laws, scale free networks and genome biology, Springer US, ISBN 978-0-387-25883-6, (2006) 1 – 11
- [12] Motter, A. E., de Moura, A. P. S., Lai Y. Ch, Dags Gupta P. : Topology of the conceptual network of language, *Phys. Rev. E* **65**, (2002) R 065102
- [13] Náther P., Markošová M., nepublikované
- [14] Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N., Barabási, A. L. : Hierarchical organization of modularity in metabolic networks, *Science* **297**, (2002) 1551 – 1555
- [15] Chialvo, D. R. : Critical brain networks, *Physica A* **340**, (2004) 756 – 766
- [16] Barabási, A. L., Oltvai Z. N. : Network biology : understanding the cell's functional organization, *Nature Reviews Genetics* **5**, (2004) 101 – 113
- [17] <http://unbound.biola.edu>

² Články možno nájsť na serveri xxx.lanl.gov pod uvedeným číslom

In silico simulácia etnických konfliktov

Jiří Pospíchal¹ a Vladimír Kvasnička²

Vyrastal som v židovskej štvrti mesta Flatbush, ktorá susedila s talianskym osídlením. Niekedy, na našej ceste do školy, niektoré talianske deti, ktoré takmer všetky chodili do cirkevnej školy, nám nadávali a dokonca nás aj napadali. Aj keď žili iba niekoľko blokov od nás, nepoznali sme ich mená. Jednoducho sme ich volali „decká zo Svätého Brennana“. Keď naši rodičia videli utržené rany, nahlásili incidenty riaditeľovi našej školy, ktorý bol Žid, ale z inej časti mesta. Ten kontaktoval relevantné autority zo školy Svätého Brennana, ktorí vyšetrili záležitosť a potrestali vinníkov. Smiešne bolo, že nikoho, ako sa zdalo, ani nenapadlo volať políciu. Boli to Íri.

Z detských spomienok Davida Laitina

Abstrakt. Evolučným algoritmom sme vyvíjali stratégie v hre „iterovaná väzenská dilema“, prispôsobenej na štúdium spolupráce medzi dvoma etnickými skupinami. Nastavenie „kolektívna vina“, kedy je potrestaná celá etnická skupina za akt nespôlupráce svojho člena voči členovi druhej etnickej skupiny, nevedlo k emergencii spolupráce medzi skupinami, dokiaľ sme nezaviedli administrátora často pokutujúceho akty nespôlupráce. Nastavenie „osobná zodpovednosť“, kedy člen skupiny, ktorý nespôlupracuje v rámci medzietnických interakcií, je potrestaný osobne v rámci svojej etnickej skupiny, už viedlo k spolupráci skupín. Simulačné výpočty boli podoporené aj teoretickou analýzou. Táto kapitola vznikla na základe našich prác [16-18].

1 Úvod

Použitie hry „väzenská dilema“ patrí v literatúre k často používaným jednoduchým prístupom k štúdiu rozdielnych pohľadov na spoluprácu tak v zvieracej ako aj v ľudskej spoločnosti. Táto hra bola obsiahle analyzovaná z veľa perspektív v biologických [20] a sociálnych vedách [1,2] a v informatike [19]. Snáď najaktívnejšia oblasť výskumu väzenskej dilemy sa týka evolučnej verzie hry. Stratégie ktoré majú nízky bodový zisk vymierajú, zatiaľ čo stratégie s vysokým bodovým ziskom sa rozmnožujú, pričom tento proces je stále opakovaný. Úspech stratégie v evolučnej hre väzenskej dilemy tak závisí viac na

¹ Fakulta informatiky a informačných technológií STU, 842 16 Bratislava, e-mail: pospichal@fiit.stuba.sk

² Fakulta informatiky a informačných technológií STU, 842 16 Bratislava, e-mail: kvasnicka@fiit.stuba.sk

dobrom skóre oproti iným úspešným stratégiám, ako na dobrom skóre voči širokému spektru stratégií. Cieľom tohto príspevku je použiť evolučnú väzenskú dilemu k štúdiu spolupráce medzi rozdielnymi etnickými skupinami, čo je v súčasnosti často diskutovaný problém. Náš záujem o tento aktuálny problém bol iniciovaný podnetnou prácou Fearona a Laitina [8], kde bola použitá štandardná teória iterovanej väzenskej dilemy na diskusiu rozdielných mechanizmov medzietnickej spolupráce a konfliktov. Dotyční autori veľmi dôkladne diskutovali ich výsledky z pohľadu rovnovážnych podmienok a dokázali veľa teorémov. My sme zmodifikovali a rozšírili ich prístup na evolučnú metódu, kde sa uvažuje o populácii jedincov. Každý jedinec je vybavený zafixovanou stratégiou väzenskej dilemy a potrestaní jedinci sú „penalizovaní“. Vhodnosť takéhoto prístupu bola v poslednej dobe podporená aj meraniami pozitronovej tomografie u ľudí pri altruistickom trestaní, ktoré naznačujú, že ľuďom trestanie podvodníkov pôsobí uspokojenie [7]. Víťazné stratégie sú diskutované a interpretované v rámci teórie evolučnej stability [3-5,21], ktorá patrí medzi veľmi efektívne teoretické prostriedky pre interpretáciu a lepšie pochopenie evolučne emergujúcich stratégií. Dôležitosťou koevolúcie kultúry s génmi pri evolúcii kooperácie sa zaoberali s inými modelmi aj Fehr a Fischbacher [9], ktorí dôležitosť penalizácie nezúčastnenou stranou modelovali vlastným modelom [10], ako aj experimentmi s kmeňmi v Papui-Novej Guineji [6].

2 Základné princípy hry „väzenská dilema“

Hra „väzenská dilema“ (prisoner’s dilemma (PD)) [20,2] je hrou medzi dvoma hráčmi, kedy každý z nich má výber z dvoch možností: spolupracovať (cooperate (*C*)) alebo podvádzať (defect (*D*)). Matica platieb tejto hry je daná ako

$$A = \begin{pmatrix} R & S \\ T & P \end{pmatrix} \Leftrightarrow \begin{pmatrix} 3 & 0 \\ 5 & 1 \end{pmatrix} \quad (1)$$

kde vstupy sú obmedzené nasledujúcimi dvoma nerovnosťami

$$T > R > P > S, \quad 2R > T + S \quad (2)$$

Keď obidvaja hráči spolupracujú, potom dostanú odmenu *R* (reward), ktorá by mala byť väčšia ako bodový zisk „trest“ *P* (punishment) získaný obidvoma hráčmi, keď podvádžajú. Ale keď jeden hráč podvádza (hrá *D*) zatiaľ čo ten druhý spolupracuje (hrá *C*), potom podvádžajúci získa výplatu bodov *T* (temptation - pokušenia) ktorá je väčšia ako *R*, zatiaľ čo spolupracujúci dostáva počet bodov *S* (sucker – ten, čo naletel) ktorá je menšia ako *P*. Druhá nerovnosť (2) znamená že celková výplata pre dvoch hráčov je väčšia keď obaja spolupracujú

v porovnaní so súčtom bodov keď jeden spolupracuje a druhý nie. Formálne môže byť matica výplat chápaná ako funkcia platieb

$$f: \{C, D\}^2 \rightarrow \{0, 1, 3, 5\} \quad (3a)$$

určená ako

$$f(C, C)=3, f(C, D)=0, f(D, C)=5, \text{ a } f(D, D)=1 \quad (3b)$$

Hra PD je často používaná v literatúre [14] pre simuláciu spolupráce v biológii, sociálnych vedách, ekonómii, atd. V prezentovanom príspevku bude PD zväčša používaná vo svojej evolučnej podobe [1,2] pripomínajúcej multiagentové systémy, ktoré v súčasnosti patria k základným konceptom bádania v oblasti umelého života. Výsledky našej simulácie budú používať teóriu evolučnej stability, pôvodne vypracovanú s biologickými aplikáciami PD hry Maynardom Smithom [20,21]. Recipročný altruizmus bol preukázaný aj u opíc [13].

3 Vnútroetnická spolupráca

Účelom tejto kapitoly je načrtnúť základné črty evolučnej PD hry, ktorá bude použitá pre štúdium vzniku spolupráce vo vnútri jednej etnickej skupiny. Používame inú klasifikačnú schému jednotlivých hráčov (jedincov, agentov) ako Axelrod [1,2]. Ten postuloval, že agenti poznajú krátku predchádzajúcu históriu tak svoju, ako aj všetkých ostatných agentov. Potom stratégia agenta závisí na predchádzajúcich ťahoch oponenta, teda či spolupracoval alebo nie v niekoľkých predchádzajúcich ťahoch s ostatnými agentmi. Tu používame celkom odlišný prístup k vyjadreniu histórie interagujúcich agentov. Podľa Fearona a Laitina [8] každý agent je (alebo nie) je penalizovaný, čo pre ostatných agentov predstavuje informáciu, že daný agent nespupracoval. Penalizácia hrá zásadnú rolu v špecifikácii interakcie medzi dvoma agentmi. Penalizovanie agenta závisí na tom, či on sám a tiež aj druhý interagujúci agent už sú (alebo nie sú) penalizovaní.

Študujme populáciu P , ktorá sa skladá z členov etnickej skupiny A

$$P=A=\{a_1, a_2, \dots, a_m\} \quad (4)$$

Ako už bolo zmienené, každý agent je špecifikovaný penalizáciou, ktorá je reprezentovaná nezáporným celým číslom

$$0 \leq p(a) \leq p_{max} \quad (5)$$

kde nulové p označuje, že agent nie je penalizovaný, zatiaľ čo kladné p označuje penalizáciu a súčasne časové obdobie, ako dlho ešte bude daný agent pokladaný za penalizovaného; p_{max} je maximálny časový úsek penalizácie. Interakcia medzi dvoma agentmi je plne určená skutočnosťou, či daný agent je alebo nie je penalizovaný a vektormi stratégie. Stratégia agenta $a \in P$ je určená štvordimenziálnym vektorom

$$s(a)=(s_1, s_2, s_3, s_4) \in \{C, D\}^4 \quad (6)$$

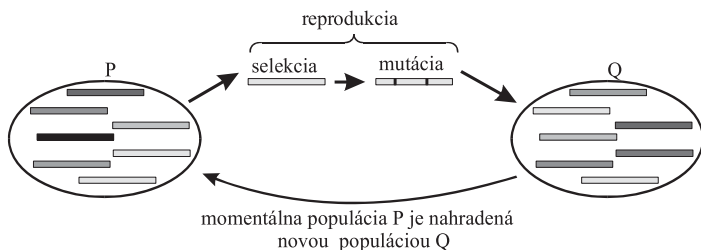
Tabuľka 1. Špecifikácia vektora stratégie $s(a)$

	penalizácia agenta a	penalizácia agenta b	ťah agenta a
1	$p(a)=0$	$p(b)=0$	$s_1(a)$
2	$p(a)=0$	$p(b)>0$	$s_2(a)$
3	$p(a)>0$	$p(b)=0$	$s_3(a)$
4	$p(a)>0$	$p(b)>0$	$s_4(a)$

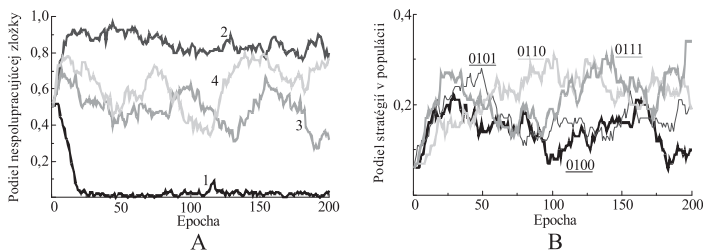
Význam jednotlivých elementov vektora stratégie $s(a)$ pre rôzne prípady je určený Tabuľkou 1, kde v poslednom stĺpci sú uvedené typy interakcie (C alebo D), ktoré používa hráč a . Napríklad, keď obidvaja agenti $a, b \in P$ nie sú penalizovaní, $p(a)=p(b)=0$, podľa prvého riadku tabuľky agent a použije ťah $move_a=s_1(a)$, podobnú tabuľku používa aj druhý hráč b , avšak v opačnom poradí druhého a tretieho stĺpca. Formálne môžeme povedať, že agent $a \in P$ je reprezentovaný usporiadanou dvojicou vektora stratégie a penalizáciou

$$a=(s,p) \quad (7)$$

Interakcia medzi dvoma agentmi $a, b \in P$ je plne špecifikovaná ich reprezentáciou (7), teda odpovedajúcou stratégiou a penalizáciou.



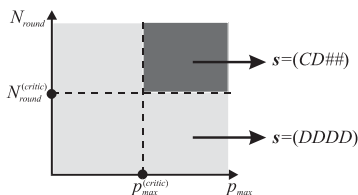
Obrázok 1. Schematické zobrazenie použitého evolučného algoritmu. Z momentálnej populácie P kvázinahodne vyberáme stratégiu do reprodukčného procesu, pravdepodobnosť jej výberu je úmerná fitness (bodovému zisku) dotyčného agenta. Reprodukčný proces spočíva v jednoduchšej mutácii vybranej stratégie, ktorej jednotlivé prvky (C alebo D) sú postupne mutované s pravdepodobnosťou P_{mut} (rovnej malému kladnému číslu, vo všetkých našich simuláciách sme použili $P_{mut}=0.001$). Výsledná nová stratégia je presunutá do novej populácie Q . Keď je veľkosť novej populácie rovná veľkosti súčasnej populácie $|P|=|Q|$, potom nová populácia nahradí tú súčasnú. Tento proces sa opakuje alebo predpísaný počet epoch, alebo až dovtedy, dokiaľ nie je výsledná populácia prakticky zložená z rovnakých alebo príbuzných stratégií.



Obrázok 2. Diagram A ukazuje grafy podielov nespolpracujúceho ťahu pre jednotlivé prvky vektorov stratégie spriemerované cez celú populáciu. Je vidno, že podiel pre prvý prvok stratégie rýchlo klesá k nule, čo znamená, že takmer všetky stratégie spolupracujú, keď obidvaja interagujúci agenti nie sú penalizovaní. Podobne druhý prvok vektora stratégie rýchlo dosiahol hodnotu blízku jednotke. To znamená, že takmer všetky stratégie nespolpracujú vo vzájomných interakciách, keď je agent dotýčnej stratégie bez penalizácie, zatiaľ čo ten druhý je označený penalizovaný. Priemery podielov nespolpracujúceho ťahu pre tretie a štvrté prvky vektorov silno fluktuujú od 0.4 do 0.8. To znamená, že sa nedá povedať, či „priemerná“ stratégia spolupracuje alebo nie, keď je jej agent penalizovaný, nech už ten druhý je alebo nie je penalizovaný. Aby sme sumarizovali dosiahnuté výsledky, môžeme povedať, že po 50 epochách je populácia stratégií takmer celkom zložená zo stratégií reprezentovaných schémou ($CD##$), kde „hash“ symboly môžu byť nahradené aj s C aj s D . Diagram B ukazuje závery z predchádzajúceho diagramu A, kde všetky ostatné stratégie okrem tých reprezentovaných pomocou schémy ($CD##$) obsadzujú zanedbateľne malý podiel populácie.

Evolúcia populácie (zloženej z chromozómov ktoré sú špecifikované podľa (7) stratégiou a hodnotou penalizácie) je simulovaná jednoduchou verziou evolučného algoritmu [15,12,11], kde je vynechaná operácia kríženia, pozri Obr. 1. Tento algoritmus je založený na proporcionálnom kvázináhodnom výbere stratégií pre reprodukčný proces. Analogicky k biológii hovoríme, že vektor stratégie v špecifikácii agenta odpovedá agentovmu genotypu, zatiaľ čo penalizačná časť odpovedá nadobudnutej vlastnosti, ktorá nie je zdedená v rámci reprodukčného procesu.

Populácia stratégií je inicializovaná náhodne, pričom každý agent má nulovú penalizáciu. Fitness agentov je vypočítané tak, že usporiadame „turnaj“, kde predpísaný počet-krát N_{round} (vo všetkých našich simuláciách používame $N_{round}=m \times m$, kde m je veľkosť populácie) náhodne vybraná dvojica agentov zohrá jedno kolo PD hry, kde ťah obidvoch je plne určený ich stratégiou a penalizáciou. Získané body sú pripočítané k fitness odpovedajúcich agentov.



Obrázok 3. Diagramatické zobrazenie oblasti príťažlivosti v závislosti na maximálnej dĺžke trvania penalizácie p_{max} a na hodnotách frekvencie interakcií agentov v každej evolučnej epoche. Slabo tienená oblasť odpovedá oblastiam parametrov produkujúcich nespolpracujúce stratégie $s=(DDDD)$, zatiaľ čo tmavšie vytieňovaný štvorec odpovedá hodnotám parametrov (obidve hodnoty väčšie alebo rovné kritickej hodnote) ktoré produkujú spolupracujúce stratégie odpovedajúce schéme $s=(CD##)$.

Numerické výsledky našich simulačných výpočtov (pre veľkosť populácie $m=1000$ a maximálne trvanie penalizácie $p_{max}=10$) sú zobrazené na Obr. 2. Vidíme, že emergujúce stratégie odpovedajú schéme \blacklozenge (CD##), ktorá (1) spolupracuje keď obidvaja agenti nie sú penalizovaní, a (2) nespolpracuje, pokiaľ daný agent nie je penalizovaný, zatiaľ čo druhý agent je penalizovaný (pozri prvé dva riadky Tabuľky 1). Víťazné stratégie pripomínajú slávnu Axelrodovu stratégiu [1,2] nazývanú „tit-for-tat“ (TFT). Naše pozorovania získané z numerických výsledkov sú zhrnuté nasledujúco:

- (1) Vývoj populácie je riadený dvoma parametrami: Kritickou frekvenciou interakcie agentov a kritickým maximálnym trvaním penalizácie. Prvý parameter určuje frekvenciu interakcií agentov, keď je menšia ako kritická hodnota, $0 \leq N_{round} < N_{round}^{(critic)}$, potom evolúcia populácie vždy produkuje nespolpracujúcu stratégiu $s=(DDDD)$. Na druhej strane, keď frekvencia interakcií agentov je väčšia alebo rovná kritickej hodnote, potom je evolúcia riadená dĺžkou trvania penalizácie. Keď je maximálna dĺžka trvania penalizácie menšia ako kritická hodnota, $0 \leq p_{max} < p_{max}^{(critic)}$, potom podobne ako v predchádzajúcom prípade evolúcia produkuje iba nespolpracujúcu stratégiu $s=(DDDD)$; keď maximálna dĺžka trvania penalizácie spĺňa $p_{max} \geq p_{max}^{(critic)}$, potom dostávame triedu stratégií, ktoré odpovedajú schéme $s=(CD##)$. To znamená, že keď je počet interakcií agenta väčší ako kritická hodnota, potom sa evolúcia populácie štiepi na dve rozdielne vetve. Prvá odpovedá víťazstvu nespolpracujúcej stratégie $s=(DDDD)$, zatiaľ čo

- Používame terminológiu z genetických algoritmov [12], kde schéma znamená množinu chromozómov, ktorých určitá časť je už špecifikovaná symbolmi C a D, zatiaľ čo zbývajúca časť (určená symbolmi #) môže mať ľubovoľné hodnoty C a D.
- Volný preklad anglického „tit-for-tat“ je „ako ty mne, tak ja tebe“.

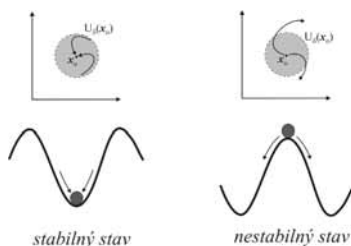
tá druhá odpovedá rozšíreniu spolupracujúcich stratégií odpovedajúcich schéme $s'=(CD\#\#)$ (pozri Obr. 3).

- (1) Pre ľubovoľnú hodnotu maximálnej dĺžky trvania penalizácie p_{max} je stratégia $s=(DDDD)$ evolučne stabilná (pozri sekciu 4). Veľkosť oblasti príťažlivosti stratégie $s=(DDDD)$ silne závisí na hodnotách frekvencie interakcii agentov a na maximálnej dĺžke trvania penalizácie p_{max} . Pre malé hodnoty obidvoch parametrov je oblasť príťažlivosti stratégie $s=(DDDD)$ veľmi veľká a môže zahŕňať všetky stratégie. To znamená, že v našich simuláciách by sa nespolupracujúca stratégia $s=(DDDD)$ mala často objavovať ako prevažujúca stratégia na konci evolúcie. Na druhej strane, keď sú veľkosti obidvoch riadiacich parametrov dostatočne veľké (teda $N_{round} \geq N_{round}^{(critic)}$ a $p_{max} \geq p_{max}^{(critic)}$), potom je veľkosť príťažlivosti $s=(DDDD)$ podstatne znížená a simulácie začnú produkovať spolupracujúce stratégie odpovedajúce schéme $s'=(CD\#\#)$.

4 Evolučne stabilné stratégie

Termín „evolučne stabilná stratégia“ bol pôvodne zavedený Maynardom Smithom [20], (pozri tiež [3-5,14,21]) v rámci jeho biologicky orientovaných aplikácií teórie evolučných hier.

Pre lepšie pochopenie pojmu evolučnej stability budeme vychádzať zo štandardnej teórie dynamických systémov, kde existuje pojem asymptoticky stabilný resp. asymptoticky nestabilný bod s jednoduchou fyzikálnou interpretáciou, pozri Obr. 4.



Obrázok 4. Stabilný stav dynamického systému je názorne reprezentovaný guľičkou v jamke. Malé vychýlenie systému z rovnovážnej polohy vyvolá procesy, ktoré ho vrátia do pôvodnej rovnovážnej polohy. V hornej časti obrázku je táto verbálna formulácia reprezentovaná pomocou trajektórie znázorňujúcej evolúciu systému. Ak počiatočný bod trajektórie je vybraný z blízkeho okolia rovnovážneho stavu, potom trajektória vždy končí v rovnovážnom bode. Pre nestabilný stav platí, že už malá výchylka stavu z rovnovážnej polohy vyvolá procesy, ktoré vedú k tomu, že systém nenávratne opustí rovnovážny stav. Potom, ak počiatočný bod trajektórie je vybraný z blízkeho okolia rovnovážneho stavu, potom trajektória nenávratne opúšťa rovnovážne okolie.

Nech populácia P obsahuje v prevažnej miere stratégiu s_0 a v malej miere aj inú tzv. votleckú stratégiu s_1 .

$$P = P_0 \cup P_1$$

$$P_0 = \{s_0, s_0, \dots, s_0\}, |P_0| = p$$

$$P_1 = \{s_1, s_1, \dots, s_1\}, |P_1| = q$$
(8)

pričom predpokladáme, že počet predstaviteľov stratégie s_0 je podstatne väčší, ako počet predstaviteľov „votleckej“ stratégie s_1 , t.j. $p \gg q \gg 0$. Stratégia s_0 sa nazýva *evolučne stabilná*, ak v priebehu evolúcie predstaviteľa inej stratégie vymiznú. V opačnom prípade, stratégia s_0 sa nazýva *evolučne nestabilná*, ak v priebehu evolúcie dochádza k zániku stratégie s_0 .

Sila stratégií z populácie P je určená vzájomným turnajom medzi jednotlivými stratégiami

$$\text{fitness}(s_0) = (p-1)f(s_0, s_0) + qf(s_0, s_1)$$

$$\text{fitness}(s_1) = (q-1)f(s_1, s_1) + pf(s_1, s_0)$$
(9)

Pretože evolúcia populácie prebieha na základe prirodzeného výberu, podmienka evolučnej stability stratégie s_0 má tvar

$$\text{fitness}(s_0) > \text{fitness}(s_1)$$
(10a)

$$(p-1)f(s_0, s_0) + qf(s_0, s_1) > (q-1)f(s_1, s_1) + pf(s_1, s_0)$$
(10b)

Z predpokladu $p \gg q \gg 0$ vyplývajú tieto dve podmienky pre evolučnú stabilitu stratégie s_0

$$f(s_0, s_0) > f(s_1, s_1), \text{ alebo} \quad (11a)$$

$$f(s_0, s_0) = f(s_1, s_1) \text{ a } f(s_0, s_1) > f(s_1, s_0) \quad (11b)$$

Týmto sme dostali jednoduché podmienky, ktoré špecifikujú či daná stratégia s_0 je v prítomnosti malého počtu votleckých stratégií s_1 evolučne stabilná. Podmienky požadujú len znalosť funkcie platieb pre jednotlivé stavy penalizácie (pozri Tabuľku 1). V našich teoretických úvahách sme použili vyššie naznačený jednoduchý formalizmus štúdia evolučne stabilných stratégií k určeniu, či výsledná stratégia je evolučne stabilná, a ak áno, za akých podmienok. Tieto štúdie evolučnej stability výsledných stratégií podstatne zvyšujú váhu našich simulačných výpočtov.

5 Princíp kolektívnej viny a medzietnickej interakcie

Cieľom tejto kapitoly je použiť prístupy z kapitoly 3 k štúdiu medzietnickej spolupráce. Riadiacim pravidlom tejto štúdie bude tzv. *princíp kolektívnej viny*, založený na predpoklade, že celá etnická skupina je zodpovedná za nespoluprácu jej člena v rámci medzietnických interakcií.

Predpokladajme, že populácia P je zložená z dvoch etnických skupín A a B

$$P = A \cup B (A \cap B = \emptyset) \quad (12a)$$

$$A = \{a_1, a_2, \dots, a_m\}, B = \{b_1, b_2, \dots, b_n\} \quad (12b)$$

Medzi agentmi populácie P existujú tri typy interakcií:

- (1) vnútroetnické interakcie v rámci etnickej skupiny A ,
- (2) vnútroetnické interakcie v rámci etnickej skupiny B , a konečne,
- (3) medzietnické interakcie medzi agentmi z A a B .

S odvolaním na Fearona and Laitina [8] postulujeme, že frekvencia medzietnickej interakcie je oveľa menšia ako frekvencia vnútroetnických interakcií v A alebo B . Táto požiadavka bude jednoducho vynútená v našich výpočtoch tým, že zavedieme pravdepodobnosť P_{inter} medzietnických interakcií v rámci množiny všetkých možných interakcií medzi agentmi z celej populácie P . Hodnota tejto pravdepodobnosti bude zadaná ako $P_{inter}=0.1$, teda iba 10% zlomok všetkých interakcií má medzietnický charakter.

Druhou základnou požiadavkou našich výpočtov je predpoklad tzv. informačnej asymetrie, definovaný v [8]. Predpokladá sa, že agenti môžu identifikovať ostatných agentov z tej istej etnickej skupiny a poznajú si navzájom svoju hernú históriu. Na druhej strane, agenti v rámci medzietnických interakcií poznajú iba to, že interagujú s niekým z tej druhej etnickej skupiny, nedokážu rozpoznať individuálnu totožnosť alebo osobnú medzietnickú históriu toho druhého agenta. Táto požiadavka sa nesnaží charakterizovať všetky typy medzietnických interakcií, napr. dlhodobé vzťahy medzi dvoma obchodníkmi. Skôr ale odpovedá špeciálnemu typu interakcií, ktoré tvoria obťažný problém pre medzietnickú spoluprácu kvôli informačnej asymetrii a relatívne nízkej frekvencii. Musíme zdôrazniť, že princíp kolektívnej viny je úzko spojený s hore uvedeným predpokladom informačnej asymetrie. Nedostatok informácií o agentoch z druhej etnickej skupiny napríklad spôsobuje, že princíp kolektívnej viny je použitý ako riadiace pravidlo pre penalizovanie celej skupiny kvôli medzietnickej nespolupráci.

Nech každý z agentov $a \in A$ a/alebo $b \in B$ je špecifikovaný penalizáciou

$$0 \leq p(a) \leq p_{max} \quad (13a)$$

$$0 \leq p(b) \leq p_{max} \quad (13b)$$

Celoetnická penalizácia je braná z intervalu

$$0 \leq p(A) \leq p_{max} \quad (14a)$$

$$0 \leq p(B) \leq p_{max} \quad (14b)$$

kde toto označenie hrá dôležitú úlohu v medzietnických interakciách ako vyjadrenie nášho predpokladu informačnej asymetrie. Pretože každý z agentov z medzietnického páru nevie nič o svojom „spoluhráčovi“, tento nedostatok informácie je nahradený oveľa slabšou požiadavkou, že interagujúci agenti poznajú, či etnická skupina ich oponentov je penalizovaná alebo nie.

Tabuľka 2. Špecifikácia ďalších štyroch prvkov vektora stratégie

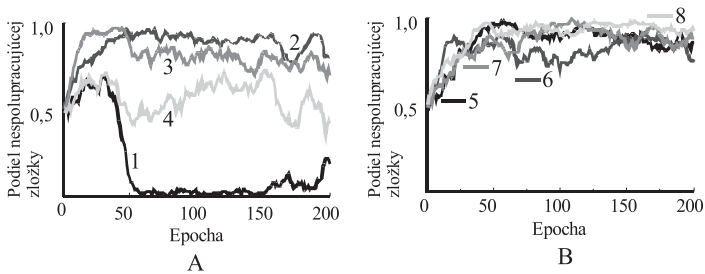
	penalizácia skupiny A	penalizácia skupiny B	ťah agenta $a \in A$
5	$p(A)=0$	$p(B)=0$	$s_5(a)$
6	$p(A)=0$	$p(B)>0$	$s_6(a)$
7	$p(A)>0$	$p(B)=0$	$s_7(a)$
8	$p(A)>0$	$p(B)>0$	$s_8(a)$

Stratégia agenta $a \in A$ je opísaná 8-dimenzionálnym vektorom

$$s(a)=(s_1, s_2, \dots, s_8) \in \{C, D\}^8 \quad (15)$$

Prvé štyri prvky vektora stratégie určujú vnútroetnickú interakciu (pozri Tabuľka 1), ďalšie štyri prvky určujú stratégiu pre medzietnickú interakciu agenta $a \in A$ s ďalším agentom $b \in B$, pozri Tabuľku 2. Evolúcia populácie (pozri Obr. 1) je robená nezávisle pre obidve etnické skupiny A a B . Etnické skupiny sú relatívne nezávislé v priebehu evolúcie, interagujú iba vtedy, keď je treba vypočítať fitness agentov, kde časť bodového ohodnotenia je určená medzietnickými interakciami.

Podobným spôsobom ako v sekcii 3 je fitness agentov vypočítaná pomocou „turnaja“, kde je predpísaný počet-krát (N_{round} , vo všetkých našich simuláciách definujeme toto číslo ako $N_{round}=m \times m + n \times n + m \times n$, kde m a n sú veľkosti etnických skupín A a B) náhodne vybraná dvojica agentov, ktorí spolu hrajú jeden ťah PD hry. Použitý ťah každého z hráčov- agentov je vždy plne určený ich vektorom stratégie a penalizáciou. Medzietnické interakcie sa v turnaji vyskytujú s pravdepodobnosťou P_{inter} , zatiaľ čo vnútroetnické interakcie sú robené s pravdepodobnosťou $1-P_{inter}$.



Obrázok 5. Diagramy A a B ukazujú grafy priemerných podielov nespolpracujúcich ťahov pre jednotlivé prvky vektorov stratégií. Prvky 1-4 sú ukázané na diagramu A, prvky 5-8 na diagramu B. Vidíme, že výsledné stratégie môžu byť reprezentované schémou $s=(CD##DDDD)$. Jasný výsledok (C alebo D) nevychádza pre štvrtý prvok vektora stratégií, často je hodnota zlomku okolo $1/2$, teda C aj D majú rovnakú pravdepodobnosť výskytu. Tento priebeh môže byť vysvetlený malou pravdepodobnosťou interakcie dvoch penalizovaných agentov, takže mutácia môže meniť podiel $D/(C+D)$ na štvrtej pozícii náhodne bez väčších následkov na fitness stratégie. Výsledná stratégia znamená, že agenti spolupracujú v rámci vnútroetnickej interakcie, keď sú obidvaja agenti nepenalizovaní, ale nikdy nespolpracujú pri medzietnickej interakcii nezávisle na tom, či sú alebo nie sú penalizovaní.

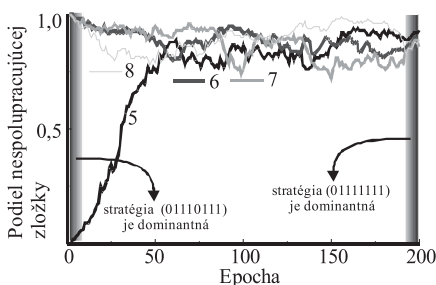
Numerické výsledky našich simulačných výpočtov (pre skupiny veľkostí $m=n=500$, s maximálnou dobou trvania penalizácie $p_{\max}=10$, a s $P_{inter}=0.1$) sú prezentované na Obr. 5 a 6. Z výsledkov sa dá usúdiť, že evolúcia stratégií, ktoré sú založené na princípe kolektívnej viny, neprodukuje evolučne stabilnú stratégiu, ktorá by mala zabudovanú spoluprácu medzi agentmi z rozdielnych etnických skupín, ktoré nie sú penalizované, pozri Obr. 6. Aj keď výsledná evolučne stabilná stratégia je spolupracujúca v rámci vnútroetnických interakcií dvoch nepenalizovaných agentov, hlavným výsledkom našich výpočtov je, že agenti z rozdielnych skupín nespolpracujú ani keď sú obidve skupiny momentálne nepenalizované. Výsledky sa dajú zhrnúť nasledovne:

- (1) Keď používame pre naše simulačné výpočty pre dve etnické skupiny rovnaké parametre ako sme použili pre výpočty pre jednu etnickú skupinu, potom výsledné stratégie odpovedajú schéme $s_0=(CD##DDDD)$. To znamená, že v rámci jednej etnickej skupiny sa vytvorí spolupráca, zatiaľ čo medzietnické interakcie medzi skupinami vyúsťujú iba v nespolpracú. Tento výsledok bol tiež overený modelovým výpočtom, kedy bola východzia populácia zložená čisto zo stratégie $s_1=(CDDDCDDDD)$, pozri Obr. 6. Stratégia spolupracujúca pri medzietnických interakciách bola rýchlo nahradená stratégiami, ktoré spadajú pod medzietnicku nespolpracujúcu schému $s_0=(CD##DDDD)$. Inak povedané, spo-

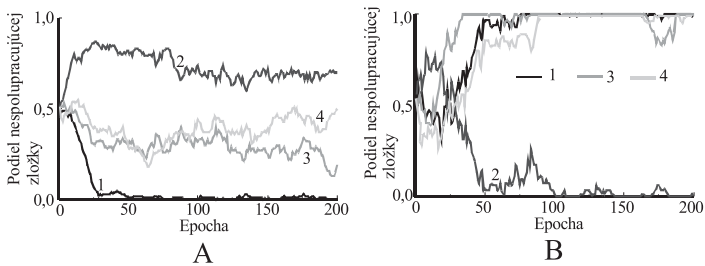
lupracujúca stratégia $s_1=(CDDDCDDD)$ nie je evolučne stabilná, teda princíp kolektívnej viny nevedie k spolupráci medzi rozdielnymi etnickými skupinami. V ďalšej časti sekcie používame formalizmus navrhnutý v sekcii 4.

- (2) Voľne povedané, princíp kolektívnej viny nie je schopný garantovať vznik spolupráce medzi dvoma etnickými skupinami. Kolektívne penalizovanie nie je dostatočne špecifické, medzietnická nespôlupráca sa šíri celou etnickou skupinou. Penalizácia celých etnických skupín preto nemôže byť použité ako hnacia sila pre vznik medzietnickej spolupráce.

V poslednej časti tejto sekcie budeme oddelene študovať prípad medzietnických interakcií, kedy sú veľkosti etnických skupín podstatne odlišné, $|A| \gg |B|$, teda počet členov A je oveľa väčší ako počet členov B . Napríklad, keď $|A|=1000$ a $|B|=100$, a počet medzietnických interakcií za epochu je 50, potom pravdepodobnosť takej interakcie pre agenta $a \in A$ za epochu je $p_A=0.05$, zatiaľ čo pre agenta $b \in B$ je to $p_B=0.5$. Z toho sa dá vyvodzovať, že pre malú etnickú skupinu B môžu byť medzietnické interakcie podstatne významnejšie ako pre veľkú etnickú skupinu A . Aby sme overili túto hypotézu, zopakovali sme predchádzajúce výpočty pre $|A|=1000$ a $|B|=100$, kde sme dali $N_{round}=10^6$, maximálnu dĺžku trvania penalizácie $p_{max}=10$ a pravdepodobnosť medzietnickej interakcie $P_{inter}=0.1$ (teda iba 10% interakcií z N_{round} je medzietnickeho charakteru). Diagramy na Obr. 7 zobrazujú dosiahnuté numerické výsledky. Vidíme, že princíp kolektívnej viny pre menšie etnické skupiny je devastujúci, pretože má vážne dôsledky aj pre vnútroetnickú spoluprácu, ktorá pre dlhé časové obdobia zaniká.



Obrázok 6. Diagramatická prezentácia našich modelových výpočtov, kde bola populácia P inicializovaná plne stratégiou $(CDDDCDDD)$. To znamená, že obidvaja interagujúci agenti, ktorí nie sú penalizovaní (alebo ktorých etnické skupiny nie sú penalizované) spolupracujú, vo všetkých ostatných prípadoch nastáva nespôlupráca. Vidíme, že táto stratégia nie je stabilná, v priebehu evolučného procesu je spontánne transformovaná na stratégiu $(CDDDDDDD)$.



Obrázok 7. Grafy podielov nespolupracujúcich ťahov pre prvé štyri prvky vektora stratégií pre etnické skupiny *A* a *B* ktoré sú rozdielnej veľkosti, $|A| \gg |B|$. Diagram A (etnická skupina *A*) ukazuje, že emergujúca stratégia je vnútroetnicky spolupracujúca pre prípad, že žiaden z dvojice interagujúcich agentov nie je penalizovaný. Na druhej strane diagram B (menšia etnická skupina *B*) ukazuje, že emergujúca stratégia je vnútroetnicky nespolupracujúca, keď sú obidvaja agenti nepoznačení.

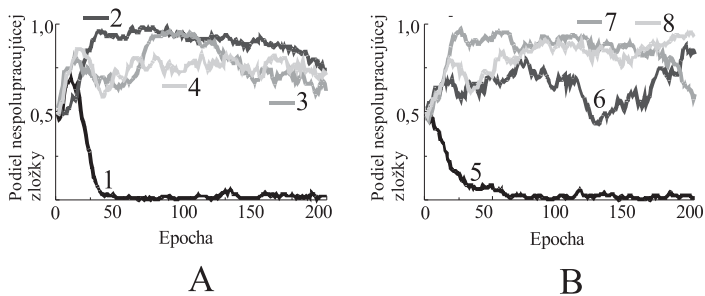
Na obrázku 8 graf A ukazuje, že pre prvú skupinu podiel nespolupracujúcich ťahov v populácii pre prvý prvok vektora stratégie sa blíži k 0, čo znamená spoluprácu nepoznačených agentov a druhý vstup sa blíži k 1, čo znamená nespuprácu poznačených agentov z *A*. Hodnoty pre ostatné dva prvky vektora sa pohybujú stochasticky okolo 0.5. Pretože pravdepodobnosť interakcie dvoch poznačených agentov z *A* je veľmi malá, náhodné mutácie zničia akúkoľvek pravidelnosť. Ostatný graf B pre vnútroetnické interakcie vykazuje zánik spolupráce u nepoznačených agentov, a napodiv spoluprácu nepoznačených agentov s poznačenými. Tento výsledok je ale iba dočasný, stratégia sa v ďalších generáciách preklápa tam aj naspäť na stratégiu (CDDDDDDD). Táto nestabilita je spôsobená genetickým driftom, pretože populácia *B* je desaťkrát menšia ako *A*, a druhá časť chromozómu pre medzietnickú interakciu ovplyvňuje *B* oveľa viac ako *A*, pretože medzietnické interakcie sa vyskytujú u *B* desaťkrát častejšie ako *A*.

6 Princíp kolektívnej viny s administrátorom

V sekcii 5 sme pozorovali, že princíp kolektívnej viny aplikovaný na medzietnické interakcie nestačí na podporu vzniku medzietnickej spolupráce. Cieľom tejto sekcie je jednoduchá modifikácia tohto princípu zavedením administrátora, ktorý trestá nespuprácu pokutou.

Modifikácia princípu kolektívnej viny zavedením administrátora bola pôvodne diskutovaná Fearonom a Laitinom [8] ako alternatívny prístup, ktorý je schopný zabezpečiť medzietnickú spoluprácu v modeli založenom na princípe kolektívnej viny. Administrátor trestá stochasticky pokutou agentov, ktorí nespupracujú v rámci medzietnic-

kých interakcií (keď obidve etnické skupiny nie sú penalizované). Podľa našich základných predpokladov o informačnej asymetrii je rozumné predpokladať, že administrátor nepozná detailné informácie o práve interagujúcich agentoch z rozdielnych etnických skupín. Predpokladajme, že agent $a \in A$ nespoupracuje s agentom $b \in B$. Potom administrátor v štýle *deus ex machina* (napr. člena Kráľovskej kanadskej jazdnej polície) stochasticky penalizuje nespouprácu v medzietnickej interakcii, pričom obidve skupiny sú v nepenalizovanom stave a druhý agent spoupracuje.



Obrázok 8. Grafy podielov nespoupráce a spoupráce pre jednotlivé prvky vektora stratégií pre vnútroetnické (A) a medzietnické (B) interakcie, kde bol zavedený dohľad administrátora s pravdepodobnosťou pokutovania $P_{penal}=0.5$. Obidva grafy, ktoré odpovedajú etnickej skupine A (pre skupinu B dostávame obdobné grafy), ukazujú, že vnútroetnická rovnako ako medzietnická spoupráca emergovala v prípade, že princíp kolektívnej viny je doplnený dohľadom administrátora, ktorý penalizuje medzietnickú nespouprácu.

Naše numerické výsledky ukazujú, že takáto jednoduchá modifikácia nášho modelu (nastavenie pravdepodobnosti pokuty je $P_{penal}=0.5$, teda asi 50% prípadov medzietnickej nespoupráce je pokutovaných) vedie k objaveniu sa medzietnickej spoupráce, pozri Obr. 8. Pravdepodobnosť pokutovania má hraničnú hodnotu, pod ktorou medzietnická spoupráca nenastáva.

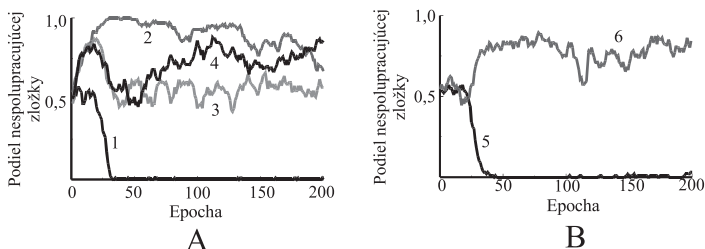
7 Princíp osobnej zodpovednosti a medzietnickej interakcie

V sekcii 5 sme rozpracovali model medzietnických interakcií založený na princípu kolektívnej viny. Aplikácia tohto princípu nevedla k medzietnickej spoupráci, a dokonca u menšej skupiny spôsobovala rozpad vnútroetnickej spoupráce. Tieto a ďalšie dôvody nás viedli k zavedeniu iného modelu medzietnických interakcií založeného na princípe osobnej zodpovednosti.

Tabuľka 3. Špecifikácia posledných dvoch prvkov vektora stratégie

	penalizácia A	ťah $a \in A$
5	$p(A)=0$	$s_5(a)$
6	$p(A)>0$	$s_6(a)$

Zameňme princíp kolektívnej viny za princíp osobnej zodpovednosti, kde sú agenti, ktorí nespolupracujú v rámci medzietnickej interakcie, potrestaní v rámci svojej vlastnej skupiny. To znamená, že na rozdiel od princípu kolektívnej viny tu existuje iba jeden druh penalizácie, kde sú agenti penalizovaní jednotlivo. Agenti inej skupiny môžu penalizovať druhého agenta, ale sami sú schopní rozpoznať penalizáciu iba u agenta z rovnakej etnickej skupiny. Agenti v rámci jednej skupiny medzi sebou nerozlišujú, či je druhý agent penalizovaný, čiže nespolupracoval s agentom z tej istej, alebo z inej etnickej skupiny.



Obrázok 9. Diagram A ukazuje priebehy pre prvé štyri prvky vektora stratégie, ktoré sú zodpovedné za vnútroetnickú interakciu. Vidíme, že prvok s_1 rýchlo ide k nule, teda prevažujúce stratégie spolupracujú pri vnútroetnických interakciách. Diagram B zobrazuje priebehy pre piaty a šiesty prvok vektora stratégie, ktoré sú zodpovedné za medzietnickú interakciu. Prvok s_5 rýchlo ide k nule, teda prevažujúce stratégie spolupracujú pri medzietnických interakciách. Emergujúce stratégie môžu byť formálne reprezentované schémou (C### CD).

Stratégia agenta $a \in A$ je popísaná 6-rozmerným vektorom

$$s(a) = (s_1, s_2, \dots, s_6) \in \{C, D\}^6 \quad (16)$$

Prvé štyri prvky tejto stratégie určujú správanie sa v rámci vnútroetnickej interakcie (pozri Tabuľka 1), ďalšie dva prvky určujú medzietnickú interakciu agenta $a \in A$, pozri Tabuľku 3. Ťah agenta v medzietnickej interakcii je plne určený jeho vlastnou penalizáciou, označenie druhého agenta nie je brané do úvahy. Táto interpretácia medzietnických interakcií je iba dôsledkom našich predpokladov o informačnej asymetrii, kde sa predpokladá, že agenti môžu identifikovať druhých agentov iba z tej istej etnickej skupiny a že aspoň do istej miery poznajú históriu hry svojho protihráča. Na druhej strane agenti pri me-

dzietnickej interakcii vedia iba to, že hrajú s niekým z druhej etnickej skupiny.

Aplikácia princípu osobnej zodpovednosti vedie k emergencii spolupráce tak v rámci vnútroetnických, ako aj v rámci medzietnických interakcií. Diagramy na Obr. 9 zobrazujú naše typické výsledky dosiahnuté pre veľkosť skupín $m=n=500$, maximálne trvanie penalizácie $p_{max}=10$, a pravdepodobnosť medzietnickej interakcie $P_{inter}=0.1$. Naše pozorovania z numerických výpočtov sú veľmi podobné tým, ktoré boli zosumarizované v sekcii 3 venovanej vnútroetnickým interakciám. Konkrétne sme pozorovali, že sa vyskytujú dva druhy evolučne stabilných stratégií. Prvá, absolútne nespolpracujúca stratégia $s=(DDDDDD)$ je evolučne stabilná, veľkosť oblasti príťažlivosti podstatne závisí na maximálnom trvaní penalizácie p_{max} , čím je p_{max} väčšia, tým je oblasť príťažlivosti menšia. Pre dostatočne veľkú hodnotu p_{max} sa stávajú evolučne stabilnými stratégie odpovedajúce „spolpracujúcej“ schéme $s=(C###CD)$, kde tieto stratégie emergujú ako prevažujúce z náhodne inicializovanej populácie stratégií.

8 Záver

Štúdium emergencie vnútro- a medzietnickej spolupráce je v našej práci uskutočnené pomocou hry „väzenská dilema“, pričom každý agent alebo tiež aj celá etnická skupina je špecifikovaná nezáporným celým číslom reprezentujúcim tzv. penalizáciu. Študovali sme populáciu stratégií agentov, ktorí patria do jednej etnickej skupiny. Ukázali sme, že v rámci jednej etnickej skupiny emergencia spolupracujúcej stratégie závisí na dĺžke trvania penalizácie a na počtu medzietnických interakcií v priebehu jednej evolučnej epochy. Ak sú tieto parametre dostatočne malé, potom prevláda nespolpracujúca stratégia $s=(DDDD)$ a spolupracujúca stratégia $s=(CDDD)$ (analog Tit-For-Tat spolupracujúcej stratégie) neemerguje. Táto nespolpracujúca stratégia je stabilná aj pre vyššie hodnoty trvania penalizácie a vyššiu frekvenciu medzietnických interakcií, aj keď so zväčšovaním týchto hodnôt sa jej oblasť príťažlivosti zmenšuje. Novým prvkom pre tieto vyššie hodnoty frekvencie medzietnických interakcií je emergencia spolupracujúcej stratégie $s=(CDDD)$ z náhodne inicializovaných populácií stratégií. Tieto výsledky sú jednoducho vysvetlené použitím teórie evolučnej stability.

K štúdiu medzietnickej kooperácie sme použili dva rozdielne prístupy ako rozšíriť jednoduchú teóriu vnútroetnických interakcií. *Po prvé*, princíp kolektívnej viny je založený na predpoklade, že celá etnická skupina je zodpovedná za nespolpracú jedného svojho člena, teda zodpovednosť za individuálnu nespolpracú sa

rozprestríe na celú etnickú skupinu. Ukázali sme, že tento model medzietnickej interakcie nevedie k medzietnickej spolupráci. Vít'azili iba také stratégie, ktoré sú vnútroetnicky spolupracujúce, ale medzietnicky nespolocujúce. Tento nepríjemný rys princípu kolektívnej viny sa dá odstrániť, keď pre obidve etnické skupiny zavedieme administrátora (*deus ex machina*), ktorý stochasticky pokutuje agentov, ktorí nespolocovali v rámci medzietnických interakcií. Zavedenie administrátora umožňuje emergenciu medzietnickej spolupráce, ale za tento pozitívny fakt sa platí relatívne vysokou frekvenciou pokút od administrátora. Ďalej, keď je populácia zložená z etnických skupín rozdielnych veľkostí, princíp kolektívnej viny môže byť devastujúci pre malé etnické skupiny, kde môže vnútroetnická spolupráca zaniknúť alebo sa stať náhodným javom. Toto zaujímavé pozorovanie môže byť vysvetlené v rámci úvah o evolučnej stabilite tým, že jedinci menšej skupiny majú relatívne vysokú frekvenciu medzietnických interakcií, ktorá spôsobuje silne „šumové“ ohodnocovanie vnútroetnických častí stratégie. Tento „šum“ spolu s faktom, že menšie populácie sú viac náchylné ku genetickému driftu, spôsobuje dlhé obdobia prevahy nespolocujúcich stratégií vo vnútroetnických interakciách

Za druhé, princíp osobnej zodpovednosti je prirodzeným rozšírením našej teórie vnútroetnických interakcií, jedinci ktorí nspolocujú v medzietnických interakciách sú znevýhodnení nspolocujúcou v rámci svojej vlastnej skupiny. Na rozdiel od princípu kolektívnej viny v tomto prípade existuje iba jeden druh penalizácie, jedinci sú za svoju nspolocujúcu zodpovední osobne. Ukázali sme, že v rámci tohto prístupu sa objavujú medzietnicky spolupracujúce stratégie. Tento fakt sa dá vysvetliť pomocou teórie evolučnej stability.

Hlavným cieľom našich simulačných štúdií o vzniku spolupráce medzi dvoma etnickými skupinami bolo lepšie pochopenie faktorov, ktoré môžu podporiť vznik medzietnickej spolupráce alebo môžu spôsobiť jej zánik. Jednoznačne sme ukázali, že princíp kolektívnej viny nevedie k vytvoreniu spolupráce medzi dvoma skupinami, pričom pre menšiu skupinu môže mať devastujúci účinok aj pre vnútroetnickú spoluprácu. Medzietnická spolupráca sa pri použití princípu kolektívnej viny objavuje až po zavedení administrátora, ktorý stochasticky s vysokou pravdepodobnosťou pokutuje nspolocujúcu. Diametrálne odlišná situácia nastáva pri použití princípu osobnej zodpovednosti k štúdiu medzietnickej spolupráce. V tomto prístupe, kde jedinci sú už osobne zodpovední za prípadnú nspolocujúcu s jedincami z inej etnickej skupiny, nastáva spontánna emergencia evolučne stabilnej medzietnickej spolupráce

Na záver tohto príspevku obrátíme našu pozornosť znovu k úvodnému mottu od Davida Laitina, spoluautora práce, ktorou bol náš príspevok zásadným spôsobom stimulovaný. Na základe prezentovaných výsledkov môžeme konštatovať, že obe etnika, tak židovské ako aj írské, sa správali podľa modelu osobnej zodpovednosti, t.j. na prípadné medzietnické konflikty nereagovali penalizovaním celého druhého etnika. Reagovali tak, že v rámci svojho etnika penalizovali tých svojich príslušníkov, ktorí v medzietnickej interakcii boli nekooperatívni (írské deti, ktoré bili židovské deti, boli v rámci írskoho etnika potrestané - penalizované). ***Snád' táto práca bude slúžiť ako malý podporný prostriedok k všeobecnej zásade, že medzietnická kooperácia spontánne emerguje len vtedy, ak všetci príslušníci oboch etník sú osobne zodpovední za kooperatívnu medzietnickú interakciu. V tomto prípade, ten jedinec, ktorý bol nekooperatívny bude potrestaný - penalizovaný vlastným etnikom. Na základe našich simulačných výpočtov vieme, že tento evolučne stabilný mód medzietnických interakcií je schopný zabezpečiť medzietnickú evolučne stabilnú kooperáciu.***

Pod'akovanie: Práca bola podporená grantmi VEGA 1/4053/07, 1/0804/08 a APVV-20-002504.

Referencie

- [1] Axelrod, R.: *The Complexity of Cooperation. Agent-Based Models of Competition and Collaboration*. Princeton, NJ: Princeton University Press 1997.
- [2] Axelrod, R.: *The Evolution of Cooperation*. New York, NY: Basic Books 1984.
- [3] Bendor, J., Swistak, P.: Evolutionary Equilibria: Characterization Theorems and Their Implications. *Theory and Decision* **45** (1998) 99-159.
- [4] Bendor, J., Swistak, P.: The Evolutionary Stability of Cooperation. *American Political Science Review* **91** (1995) 290-307.
- [5] Bendor, J., Swistak, P.: Types of Evolutionary Stability and the Problem of Cooperation. *Proc. Natl. Acad. Sci. USA* **91** (1995) 3596-3600.
- [6] Bernhard, H., Fischbacher, U., Fehr, E.: Parochial altruism in humans. *Nature* **442**, (2006) 912-915.
- [7] De Quervain, D., Fischbacher, U., Treyer, V., Schellhammer, M., Schnyder, U., Buck, A., Fehr, E.: The neural basis of Altruistic Punishment. *Science* **305** (2004) 1254-1258.
- [8] Fearon, J. D., Laitin, D. D.: Explaining Interethnic Cooperation. *American Political Science Review* **90** (1996) 715-735.
- [9] Fehr, E., Fischbacher, U.: The nature of human altruism. *Nature* **425** (2002) 785-791.
- [10] Fehr, E., Fischbacher, U.: Third Party Punishment and Social Norms. *Evolution and Human Behavior* **25** (2004) 63-87.
- [11] Fogel, D. B.: *Evolutionary Computation. Toward a New Philosophy of Machine Intelligence*. New York, NY: The IEEE Press 1995.
- [12] Goldberg, D.E.: *Genetic Algorithms in Search, Optimization, and Machine Learning*. Reading, NA: Addison-Wesley Publishing Company 1989.
- [13] Hauser, M.D., Chen, K., Chen, F., and Chuang, E.: Give unto others: genetically unrelated cotton-top tamarin monkeys preferentially give food to those

- who give food back. *Proceedings of the Royal Society*, London, **B 270** (2003) 2363-2370.
- [14] Hofbauer, J., Sigmund, K.: *Evolutionary Games and Population Dynamics*. Cambridge, UK: Cambridge University Press 1998.
- [15] Holland, J.H.: *Adaptation in Natural and Artificial Systems*. Ann Arbor, MI: University of Michigan Press 1975.
- [16] Kvasnička, V., J. Pospíchal: *A Multi-agent Study of Interethnic Cooperation*. Lecture notes in artificial intelligence, vol. 2086, Springer 2001, pp. 415-435.
- [17] Kvasnička, V., Pospíchal, J.: Chapter 18: Evolutionary Study of Interethnic Cooperation. In: F. Schweitzer (ed.): *Modeling Complexity in Economic and Social Systems*. Singapore: World Scientific, 2002, pp. 293-322.
- [18] Kvasnička, V., Pospíchal, J.: Evolutionary Study of Interethnic Cooperation. *Advances in Complex Systems* **2**(4) (1999) 395-421.
- [19] Lindgren, K.: Evolutionary Phenomena in Simple Dynamics. In *Artificial Life II*, (Langton, C. G., Taylor C., Farmer J. D., Rasmussen, S., eds.). Redwood City, CA: Addison-Wesley Publishing Company 1992.
- [20] Maynard Smith, J.: *Evolution and the Theory of Games*. Cambridge, UK: Cambridge University Press 1982.
- [21] Weibull J. W.: *Evolutionary Game Theory*. Cambridge, MA: MIT Press 1997.

Co je to přirozený a co je to umělý

Pstružina Karel¹

Abstrakt. V této stati se chceme zamyslet nad tím, co je přirozené a co umělé. Tento problém se nám zdá důležitý právě z toho úhlu pohledu, který souvisí s problematikou možných modelů mysli a lidského myšlení. Ve stati se nejprve zamyslíme nad vymezením toho, co je přirozené a toho, co je umělé, abychom se na tomto základě pokusili o zvážení možnosti nakolik je imitace přirozeného uskutečnitelná v oblastech mysli a lidského myšlení. Stať zakončujeme úvahou o rovnováze mezi přírodním a umělým.

1 Přirozenost a umělost

Za přírodu považujeme vše, co má bezprostřední souvislost se jsoucný. Příroda je pro nás celek jsoucen (nikoli věcí). Příroda je tedy mnohem širší pojem než svět nebo skutečnost, neboť obsahuje i to, co není vnímáno a tedy i konstituováno lidským myšlením se stimulů vytržených se jsoucen, což se týká jak jednotlivých vlastností věcí, tak i vztahů mezi jsoucný. Zároveň poznamenáváme, že je značné množství toho, co nebylo a není vnímáno, ale co jestvuje a je tedy určité a tedy patří zároveň do přírody. Přirozeným je vše, co je přírodou a co je přírodou zrozené, což znamená, že také lidské myšlení je přirozené, neboť se zrodilo z přírodních dějů. Jeho průběhy (a tím i operace lidského myšlení) jsou určeny jak biologickou vazbou na centrální nervovou soustavu, tak také i pohyby mezi jsoucný při projekcích, neboť i ty omezují, nebo vytyčují jednotlivé pohyby lidského myšlení. Lidské myšlení je však ovlivněno nejen svou přirozenou a přírodní vazbou na biologický substrát, ale jeho pohyby jsou mnohdy vychylovány ze své původní dráhy zvyky, předsudky a dalšími společenskými okolnostmi, což již nemůžeme za přirozené považovat, neboť to vše jsou motivy, nepocházející z přírody, ale ze záměrů a plánů, nejsou tedy něčím bezprostředním, ale zprostředkovaným.

Souhrn věcí, tedy svět je také ještě přirozeným, neboť je spontánně konstituován lidským myšlením. Demarkační linii však přesto vedeme již zde (tzn. v oblastech konstitucí věcí), neboť některé z konstitucí jsou vázány na teoretické modely, tj. na abstrakta, čímž pohyby lidského myšlení přesunují ze spontánních konstitucí ke konstrukcím. Takovými teoretickými modely jsou např. teze: země se otáčí kolem slunce a mnohé další (patří mezi ně také budování domů a celých městských komplexů v pravoúhlých liniích). V běžných konstitucích

¹ Vysoká škola ekonomická, W. Churchilla 4, Praha 3, 130 00, e-mail: Pstruzin@vse.cz

světa lidské myšlení nezaznamenává to, že by se země otáčela kolem vlastní osy. Je zde však teoretický koncept, kterým prohlížíme a vysvětluje si svět a ten upravuje něco nepocit'ované do teoretické roviny a tak je naším objasněním to, v němž přiřkneme zeměkouli otáčení kolem vlastní osy a tím si i vysvětlíme pohyb slunce na obloze během dne. To však je již konstrukce pramenící z pohybů lidského myšlení v mysli mezi mentály. Je to již dílem, kterým následně prohlížíme a objasňujeme si svět, a který má již nějaký i když ne vždy zjevný účel.

Co tedy demarkační linie mezi přirozeným a umělým předznamenává?

Především to, že umělé je dílem umu (chtělo by se říci roz-umu), tedy intelektu. Není zde přítomná jen spontánní a bezprostřední konstituce, tedy konstituce na základě toho, co je mysli vrozené a tudíž i přirozené, ale v umělém je vždy něco, z takových pohybů lidského myšlení, kdy toto je již vedeno vnitřními souvislostmi a nikoli jen projekcemi. Umělé vyvěrá z pohybů, které jsou prováděny jen v samotné mysli a vyúsťují v konstrukcích předmětů a v tvorbě díla. Jde tedy o hierarchizované pohyby. Nejprve je projekce mezi jsoucna (přírodu) s následnou konstitucí věci a mentálu a druhým stupněm je pak pohyb lidského myšlení již jen v mysli samotné, tedy mezi reprezentacemi (mezi vjemy) a z nich tvorba něčeho, co je zrozeno z umu, tedy umělého.

Především je nutné upřesnit, že ne každý pohyb lidského myšlení v regionu mysli vyúsťuje v tvorbu něčeho umělého. Jako výsledky takovýchto pohybů mohou být např. spontánní fantazie, sny, bezděčné narace, atd. Tyto pohyby jsou orientovány jen dovnitř mysli samotné a nemají nic společného s bezprostředností projekcí a konstitucemi světa. Lidské myšlení v nich spojuje jen to, co již dříve nabylo a to nikoliv pro nějaký účel, i když mnohé z těchto fantazií, snů, atd. mohou vyúsťit v konstrukce předmětu a v tvorbu díla, což již něco umělého jest. To, proč tyto aktivity lidského myšlení nemůže vřadit k umělému je proto, že jsou bezprostředně spontánními, nic nezprostředkovávají, ale jen vyplývají z perpetuálnosti lidského myšlení.

Proč jsou tedy takové produkty lidského myšlení jako je tvorba předmětu a dílo umělými? Především proto, že je zde záměr, zacílení, které není odvoditelným ze spontaneity pohybů mezi jsoucny, nebo i v mysli samé, ale lidské myšlení tím, že má záměr překračuje hranici spontaneity a vydává se do oblastí, kdy úmyslně tvoří novosti, tvůrky. Je to tedy úmysl, či záměr, co tvoří demarkační linii. V tomto směru se shodujeme s I.M. Havlem, když vymezuje účel jako rozhraní mezi přirozeným a umělým. Dokladujeme:

Velmi obecně řečeno, abychom o nějaké věci (vlastnosti, činnosti) mohli tvrdit, že je umělá, měly by být splněny přinejmenším tři podmínky:

- 1. existuje nějaká přirozená věc, logicky připouštějící duplikaci (v našem případě je to například lidské myšlení, vnímání, rozhodování apod.),*
- 2. existuje záměr člověka (nebo týmu) vytvořit duplikát oné přirozené věci,*
- 3. došlo k provedení záměru, čili proběhl intencionální proces vedoucí od záměru k jeho realizaci (v našem případě byla dotyčná věc například implementována na počítači).²*

Proč, úmysl, záměr, rozlišují přirozené a umělé? Kromě zřejmé odpovědi: protože úmysl, zacílení není zrozeno z přírody, ale ze subjektu, je zde ještě další motiv, který spočívá v tom, že umělé se mnohdy snaží přeskupit přirozené tak, aby cíl byl dosažen, a tím narušuje přírodní procesy, přeskupuje je. Příroda je mu jen pomocným materiálem, z něhož a na němž může prosadit sebe sama. Příroda není subjektem, není tím, z čeho pochází aktivita, ale stává se jen vedlejší nutností, bez níž by sice výtvar nemohl být vytvořen, ale zároveň je příroda něčím, co je dílem překročeno. Příroda sice obsahuje dílo jako svou potencialitu, ale sama proces převedení potenciálního v aktuální, nebo také možnosti ve skutečnost nerealizuje. Tvorba díla lidským myšlením je tedy hierarchickým procesem realizace další vrstvy.

Jistě je správné argumentuje-li někdo tak, že přece je v přirozenosti lidského myšlení, že má záměr, cíl; že lidské myšlení je zde tak, že jeho součástí je tvorba, a že takto bylo zrozeno přírodou, že tedy demarkace, kterou zde vedeme, je falešnou. Tuto argumentaci respektujeme, avšak naše odpověď je takováto: jaký smysl má v tomto případě uvažovat o něčem umělém(?); a jaký by byl v tomto případě význam slova umělý? Člověk je tak, že patří k jeho přirozenosti mít záměr, mít cíl pro své činy, ale tento způsob lidského pobytu jej odlišuje od jiného, který je mu také vlastní, a sice toho, při němž jen reaguje na stimuly, kdy jeho činům nevedou předem promyšlený plán. Takováto reaktivita je rovněž přirozená, aktivita je v tomto případě vedena a vynucena jsoucný, lidská reakce a aktivita nepřekračuje přirozený rámec, ale je jeho součástí. Tak je tomu ve většině případů, kdy uspokojujeme své biologické potřeby, které nejsou modifikovány našimi ambicemi. Naopak, když je příroda podrobována plánům a ambicím, pak se již jedná o umělost.

² Havel, I.M.: Přirozené a umělé myšlení jako filosofický problém. In: <http://www.cts.cuni.cz/reports/1999/CTS-99-11.htm>

2 Imitace přirozeného

Jak je tomu tehdy, když lidské myšlení ve svých pohybech imituje přírodní procesy a tato imitace je právě oním úmyslem? Naše odpověď je: také v těchto případech se jedná o umělost.³ Přírodními procesy se lidské myšlení velmi často nechává inspirovat a dokáže je imitacemi využívat ke svým účelům. Většinou však jde zároveň i o přeformování přírodních procesů a tedy o procesy umělé.

Hegel pro označení těchto procesů používá termín: *Lest rozumu*.

*Rozum je stejně lživý, jako mocný. Lest vůbec spočívá ve zprostředkující činnosti, která nutíc předměty působit na sebe navzájem podle své povahy a vzájemně se opracovávat, aniž se do tohoto procesu bezprostředně vměšuje, přesto uskutečňuje jen svůj účel.*⁴

Hegelovo pochopení vzájemného opracování části přírody jinou částí přírody je přeformováváním přírodních procesů. Lidské myšlení pozoruje přírodu a hledá a nachází v ni mnohé, co může posloužit jeho účelům. Čím hlouběji proniká k poznání přírodních dějů, tím blíže je také k jejich pochopení a tedy i uchopení s možnostmi následné manipulace s nimi, neboť tyto již jsou nějak v jeho držbě a lze je sestavovat tak, aby samy účel uskutečnily. Lze dokonce říci, že jen velmi málo vynálezů, ulehčující lidský život je svébytným výtvorem lidského myšlení, který se nezakládá na imitacích přírodních procesů. Teprve s prosazováním konstruktivní racionality ve vědě se tento proces stává dominantním.

Zvláštní kapitolou imitací je pak snaha napodobit průběhy lidského myšlení počítačným přístrojem. Také tuto imitaci považujeme za umělou. Stejně tak jako při všech imitačních procesech je i při napodobení pohybů lidského myšlení nutné nejprve postihnout samotné operace lidského myšlení (nebo alespoň některé z nich) a tyto se pak snažit přenést do přístroje.

Jedná se tedy o imitaci operací a nikoli obsahů lidského myšlení. Obsahy jsou sice snadněji napodobitelné, avšak jen co se týká jejich formy. Imitovat intencionalitu obsahů jednotlivých prožitků a pojmů

³ Jde-li se např. o výbuch A bomby, jde o nahromadění prvků uranu v nadkritickém množství, které se v přírodě obvykle v takovéto koncentraci nevyskytuje. Pochopitelně, že ve vesmíru k takovýmto a mnohem dramatičtějším procesům dochází, a jde v těchto případech o procesy přírodní, avšak to jsou jevy, které nevznikají umělým přiblížením prvku uranu v nadkritickém množství. Samotný proces výbuchu A bomby je jistě přírodním procesem, protože uran má ty vlastnosti, které má, avšak ono přiblížení, onen úmysl způsobit výbuch je procesem umělým. Stejně tak bychom mohli posuzovat mnohé další procesy, jakými jsou např. umělé oplodnění a další genetické manipulace, šíření vln různých frekvencí za účelem přenášení hlasových nebo obrazových informací atd.

⁴ Hegel, G. W. F.: Encyklopaedie. Berlín 1840, s. 382

je mnohem složitější.⁵ Intence vždy znamená, že prožitek či pojem je o něčem a takto je intence velmi plastickou vlastností myšlenkových obsahů, je vlastností, která se neustále pozměňuje a prohlubuje a s každým dalším prožitkem mírně, nebo zásadně pozměňuje svůj význam.

Intencionalita je podle našeho názoru určena ve třech rovinách. Jde o:

- Komplexitu;
- Adekváci; a
- Kontextualitu.

Komplexitou rozumíme neustále vřazování obsahu do sémantické sítě vztahů s ostatními obsahy a to podle asociativních a analogických vazeb; dále podle vazeb logicko-syntaktických; a v neposlední řadě také podle vazeb narativních. Ve všech těchto případech lidské myšlení při každém setkání se s nějakým prožitkem doplňuje intenci prožitku a pojmu a rozšiřuje svou sémantickou síť, v níž je prožitek nebo pojem ukotven. Upřesňujícím a doplňujícím zařazením prožitku a pojmu do sémantické sítě se také mění síla vazeb mezi jednotlivými významy a prožitek a pojem jsou znovu zhodnocovány vzhledem k celé endoceptivní struktuře osobnosti, takže při dalších projekcích lidského myšlení mezi sebou budou vystupovat s jiným důrazem a tak vytrhávat stimuly podle svého očekávání a svých plánů. Jde tedy o nárůst a neustále zpřesňování toho, jaký význam prožitek má.

Adekvácií pojímáme jako zdůraznění rozhodujících sémantických vztahů, které jsou dány již zvukomalebností slov. Jestliže např. nyní píšeme o **přirozeném**, pak již v tomto slovu je zakotveno to, že rozhodujícím významem tohoto slova je něco, co je **přírodou zrozené**; jestliže dále píšeme o něčem, co je **umělé**, pak rovněž v tomto slovu je již obsažen hlavní důraz na to, že jde o něco, čemu dává vznik **um** (nebo **rozum**). Tak bychom mohli pokračovat příklady o tom, co je to stanovisko, kdy v tomto slovu zaznívá jiné slovo, a sice stanovisko, tedy místo odkud se dívám, nazírám na svět, a tudíž mám názor, atd. Adekvacie nemění význam, ale jen zdůrazňuje jeho rozhodující pojetí.

Konečně se intenci velmi silně podílí také kontext, v němž je prožitek zaznamenán nebo pojem používán. Tato problematika je důkladně probírána zejména L. Wittgensteinem v jeho Filosofických

⁵ Zde hovoříme o imitaci prožitků, nebo pojmů, i když při analýzách intencionality jde převážně o práci s pojmy a jejich denotáty. Protože však naším úkolem je vypořádat se s lidským myšlením a jeho pohyby, uvažujeme i o prožitcích a jejich intencích, tzn. o čem tyto prožitky jsou, k čemu se vztahují, protože prožitek může předcházet pojmotvorbě. J. Searle (ve své práci: *The Construction of Social Reality*. New York, Free Press 1995) to označuje jako předjazykovou formu intencionality.

zkoumáních, ale také mnohými dalšími, jako např. Lakoffem, když analyzuje použití metafor v určitých kontextech.

Pochopit význam prožitku nebo pojmu nelze jinak než tak, že je zde přítomno také pozadí, v němž prožitek nebo pojem vystupuje a prokazuje se. Pozadí vytváří specifický prostor pro umístění lidského myšlení, k němuž se vše obrací a útočí na jeho pozornost, nabízí se mu. Lidské myšlení je však obraceno k prožitku a pojmu, jeho pozornost je již orientovaná, avšak v náhlosti pohybu se může něco z pozadí zablýsknout a zlákat lidské myšlení, které si již nese ke zpracování nejen něco dominantního, ale něco dominantního společně s výskytem dalších, z pozadí se vynořujících a tacitně přítomných stimulů, nebo dalších stimulů. Při zpracování dominantního stimulu, je tento zařazován do sémantické sítě společně se stimulem z pozadí a tím se zvýrazňuje, nebo zeslabuje jeho význam i hodnocení.

Poznáme-li na čem se intencionalita zakládá, jak je komplexita a kontextualita závislá na prožitcích a jak je neustále doplňována a proměňována, pak by snad bylo možné imitovat také intencionální obsahy, jimiž lidské myšlení při operacích disponuje tím, že by se vybuodovala sémantická síť, která by mohla napomoci i rozlišení významů jednotlivých stimulů. Musíme si však uvědomit, že intence jsou vymožeností mysli pro generování interakcí, pro spouštění následného chování vůči stimulům a proto má smysl uvažovat o imitacích intencionálních obsahů mysli jen v tom případě, že by zde jestvovalo něco, co by mohlo na stimuly odpovídajícím způsobem reagovat, tedy robot, který sleduje účely.

U operací lidského myšlení jde při snaze o jejich imitace rovněž o obtíže, zejména u těch operací, které jsou již složitější, tzn. těch, kdy lidské myšlení využívá celé vzorce a scénáře pro určité situace. Takovým operacím, jakými jsou identita a negativita a případně i dalším, jakými jsou operace klasické formální logiky, či sylogismy, nic nestojí v cestě k jejich napodobení, neboť tyto operace lze proměnit v logické obvody a tak je materializovat. Velmi obtížné, jestli vůbec možné, je však imitace takových operací jakými jsou abstrakce a generalizace. Tyto operace mají svá specifika a dosud je nelze převést na operace klasické logiky a tudíž i do formy logických obvodů. Zřejmě v lidských mozcích takovéto obvody jsou, avšak nepůjde o jednoduchá propojení, jak je tomu např. u implikace, nebo disjunkce, atd. ale půjde spíše o excitace vzorců v celých populacích neuronů.

Znovu zde chceme zdůraznit, že mozek přijímá všechny stimuly v atomizovaných formách a že každý mozkový modul je zpracovává jako jednoduché, atomické. Vše je tedy nejprve rozložené. Teprve při konstitucích věcí a světa jsou tyto jednoduchosti syntetizovány do

celků a jako celky reflektovány. Proto např. generalizace a abstrakce mohou probíhat v mozku automaticky a nemusí jim předcházet žádné operace redukcí nepodstatného, tedy toho, bez čehož je věc stále toutéž věcí. Právě tak může docházet i ke generalizacím, protože určité jednoduchosti jsou společné celé řadě stimulů reprezentujících věci a tyto jsou následně shrnuty do jedné třídy a přiřazeny k určitým kategoriím. Kdybychom chtěli tyto operace lidského myšlení strojově napodobit, pak bychom patrně museli zkonstruovat zcela jiný typ chipů, a sice takový, který by nebyl založen na obvodech, ale na excitacích nějakých elementů.

Ještě složitější je situace při snaze imitovat komplexní scénáře a memetické vzorce. Jednak proto, že jde o rozsáhlé strukturní vzorce, jejichž přeformování do podoby logických obvodů by asi nebylo snadné, ne-li nemožné. Druhým důvodem je okolnost rozpoznání toho, proč určité memetické vzorce, či scénáře jsou upřednostňovány oproti jiným vzorcům či scénářům, na čem tedy spočívá jejich výběr a rozhodnutí o jejich uplatnění v určitých situacích a jakou roli zde hrají neuvědomované, případně i nevědomé procesy lidské mysli. Odpovědi může být poukaz na zkušenostní složku endoceptivní struktury osobnosti, avšak právě zde imitace ztroskotává, neboť zkušenost je ryze individuální záležitostí. Každý z nás má jiné prožitky (retence) a dokonce i stejný prožitek v jiném časovém kontextu může vyvolat jinou reakci. Při určení reakce na stimul jde vždy o souhru retencí a protence s aktuální situací, tedy něčeho osobního s přítomnou aktuální situací. Jde tedy o dva póly, které se setkají, zkříží a místa onoho setkání určují také odpověď, reakci. Aktuální situace je něčím odlišným od obvyklých okolností. Okolnost je to, co je okolo, co nás obklopuje a co je pro nás něčím známým. Situace je tu v náhlosti a bez orientace, je něčím vynořivším se proti našim plánům a očekáváním, je něčím v čem se musíme nejprve orientovat, abychom následně zvolili adekvátní odpověď. Tu hledáme zkušenostně, avšak ne vždy ji už předem v mysli máme, nebo máme a je někde zasutá a tudíž aktuálně nepřítomná. Někdy reakci tedy musíme nejprve stvořit a to navíc v náhlosti. Tento proces je proto stěžejně imitovatelný.

Přes všechny tyto obtíže lze říci, že postup imitací lidského myšlení je neobyčejně rychlý. To však, v čem se nám zdá, že je jednostranný, je dominantní akcentace logických pohybů. Podle našeho názoru jde sice o velmi významné pohyby, avšak logické vzorce zachycují jen některé z pohybů lidského myšlení. Vedle logicky určitelných pohybů, lidské myšlení využívá také pohyby, které se zakládají na celých scénářích, případně narativních skladbách určujících očekávání a plány, které lidské myšlení sleduje. UVědomujeme si, že na-

podobit tyto scénáře je neobyčejně obtížné, přesto je na tento nedostatek nutné upozorňovat.

3 Rovnováha

Další naše úvahy o přirozeném a umělém se ubírají k námětu, v němž zvažujeme nakolik je umělé v symbióze s přirozeným, či zda již umělé nepřevažuje, či dokonce nepřevládá a tak i ovládá přírodu. Přístup k posuzování tohoto poměru se zakládá čistě na dojmech a individuálních prožitcích.

Nejprve k tomu, co je to rovnováha.

Tato otázka se stává stále více neodbytnou. Její neodbytnost spočívá v její naléhavosti a v tom je také její bytnost. Dostáváme se totiž do stádia neustálého narušování rovnováhy dominujícím kritériem, které je uplatňováno na společenskou produkci a tím je **růst**. Toto kritérium, které velmi vystižně vyzvedává V. Bělohradský⁶ je tak silné, že se postupně prosazuje do všech oblastí společenského života a vyčlňuje přirozenou rovnováhu. Snad přijde po etapě dobývání a měnění světa doba, kdy budeme zažívat dobrodružství udržování homeostázy a kdy vědecký objev nebude sloužit jako nástroj změn, ale jako nástroj na udržování rovnováhy.⁷

Rovnováha znamená vážit rovně. Je to stará ramenná váha, jejíž páky jsou vodo-rovné a tím je vyvážená. Jestliže se však jedná o vážení něčeho tak složitějšího jako jsou jsoucná, tedy příroda na jedné straně a jejich imitace, tedy něco umělého na straně druhé a jestliže navíc nejsou tato závaží námi dokonale poznána, je obtížné říci, kolik ramen a tím i vážních misek by váha měla obsahovat, abychom vše zvážili. Zbývá tedy úvaha. Jedno však je jisté. Váha se neustále automaticky vyrovnává a dosahuje tak rovnováhy, avšak nikoli takové, jaká je sledovaná účely lidského myšlení, ale jak je to možné z hlediska jsoucné, tedy přírody.

Rovnováha je vždy nějaká. Jsoucná vždy reagují na jakýkoli zásah, který váhy zatíží a vyváží jen třeba tím, že ustanoví rovnováhu mezi novými dominantami, které dříve nebyly rozpoznány, nebo nebyly v centrálním postavení. Toto se z hlediska člověka může zdát katastrofou, ale takovéto hodnocení je jen lidským oceněním. Jsoucná jsou k němu lhostejná. Katastrofa (např. anihilace některého biologického druhu, třeba i člověka) je jen jednou z možností bytí jsoucné, při níž se vykáže, v jakých vztazích se jsoucné nachází vůči jiným jsoucnům. O narušení rovnováhy tedy můžeme hovořit jen buď ve

⁶ Srovnej např.: Bělohradský, V.: Prebendáři a výprosníci. Právo, dne 4.5. 2006

⁷ Srovnej: Kelemen, J.: Postmoderný stroj. Bratislava 1998, s. 115

vztahu k dřívějšímu stavu, nebo vzhledem k protencím, v nichž se lidské myšlení vztahuje ke kladeným účelům. Vyrovnávání rovnováhy, nebo její ustanovení však není okamžité. Většinou jde o dynamický proces, v němž se něco formuje jako strukturní uspořádání a v němž dochází ke kolísání výchozích hodnot. Toto kolísání je stále a projevují se v něm vlastnosti toho, jak jsoucna k sobě patří/nepatří a jak se navzájem ovlivňují.

Zdá se tedy, že k dosahování rovnováhy dochází neustále, i když ne vždy okamžitě (ale i hledisko času, resp. dějinnosti je také jen lidským). Doba potřebná k jejímu ustanovení je závislá na charakteru jsoucna, nikoli na charakteru konstituovaných věcí nebo dokonce konstruovaných předmětů. Ty jsou jen povrchovými ukazateli. Avšak právě to, jak budeme konstituovat věci a konstruovat předměty bude ramena váhy buď vychylovat, nebo vyvažovat a proto je nutné poznávat, jak jsoucna jestvují, abychom věděli kam rychle přiskočit a co na misku váhy přidat, ubrat, případně, co si můžeme dovolit konstituovat a konstruovat, aby se rovnováha příliš nenarušila a přitom bylo učiněno zadost naším účelům. Prosazovat účelovost a při tom vyváženost je dle našeho názoru jedním z rozhodujících úkolů, které stojí před vědou. Nejde tedy o neustálý růst poznatků, ale o nárůst takových poznatků, které nenarušují rovnováhu, ale přispívají k utváření rovnováhy.

Domníváme se tedy, že poměr mezi přirozeným a umělým je dosud vyvážený. Přirozené stále převládá a tím i vládne. Tento dojem se v nás ustálil proto, že vždy, když jde o podstatné a kritické momenty ve vývoji, vždy se přírodní projev jako určující a jeho pochopení naviguje další postupy lidského a umělého a to, ať již v jeho potlačení, či naopak v takovém jeho růstu, který ustavení rovnováhy napomáhá. Takto jsme navigování především v našem osobním životě, kdy přirozená předurčenost prostupuje skrze naše geny k naplnění. Jistě, je zde mnohé, co je již umělým a co pochází z našich úmyslů, je to vřazení do společenského života, ale ve všech kritických fázích, zdá se nám přírodní v nás promlouvá a rozhoduje. Již mnohokrát jsme se na to snažili poukázat při rozbořech funkce spánku a snů. Stručně zopakujeme to, že funkci spánku a snů vidíme v konsolidacích endoceptivní struktury osobnosti, která pak v bdělém stavu předchází jak naším vjemům, tak i naším rozhodnutím. My přes den mnohé podněty nabíráme a prožitky skladujeme, mnohdy jako neutříděné. A až spánek se jimi znovu zaobírá a sám podle svých vlastních kritérií něco vypouští a něco uchovává a asociuje s jinými retencemi a tak vlastně vše znovu přehodnocuje. V tomto procesu také patrně dochází, ke konsolidaci dlouhodobé paměti a zároveň i jejímu

roztřídění na paměť epizodickou, sémantickou a procedurální. Každá z těchto pamětí má patrně své vlastní mechanismy uchování a vybavování. Dodáváme, že také mnohé z každodenní zkušenosti poukazuje k tomu, že ve spánku, alespoň v jedné z jeho fází (REM fáze) může docházet k prolamování bariér algoritmického myšlení a k realizacím tvořivých procesů.

To vše jsou přírodní procesy, které stojí v základech našich úvah o rovnováze mezi přírodním a umělým. Domníváme se, že k převaze umělého nad přírodním bude docházet až na dalším stupni, tzn. tehdy, kdy přírodní vytvoří něco umělého a toto umělé bude mít samo už svůj účel, cíl, který bude nějak naplňován. Tak tomu může být nejen tehdy, kdy by lidské myšlení vytvořilo počítačový stroj, který by sám něco dalšího vypočetl (vymyslel) a tento výpočet (výmysl) realizoval, ale i tehdy, kdy lidským myšlením stvořený a realizovaný účel potřebuje další produkty, které takto můžeme označit jako účely tohoto umělého. Je pak lhostejné, zda se na jejich produkci bude podílet člověk, či samotný stroj.

Rovnováha mezi přirozeným a umělým je však velmi křehká a snadno se může stát, že nekompetentnost rozhodování o společenském vývoji převáží nad tím, co udržuje rovnováhu. Sdílíme však optimistické stanovisko a domníváme se, že po této vzrušující periodě, kterou právě prožíváme a kterou přirovnáváme k renesanci, dojde k novému pochopení dějin a tím i k jejich proměně.

Poděkování: Tato stať vznikla za podpory všech daňových poplatníků, kteří, i když mnozí nechtě, přispěli a grantové instituce pak následně jen rozdělily takto získané prostředky, mimo jiné také na projekt: Evropského sociálního fondu pro Cíl3 regionu NUTS 2 hl. m. Praha: Kognitivní informatika. Naše poděkování daňovým poplatníkům je upřímné.

Literatura

- [1] Bělohradský, V.: Prebendáři a výprosníci. *Právo*, dne 4.5. 2006
- [2] Havel, I.M.: Přirozené a umělé myšlení jako filosofický problém. In: *Umělá inteligence III*. (V. Mařík a kol., editoři), Academia, Praha 2001, s. 1775 <http://www.cts.cuni.cz/reports/1999/CTS-99-11.htm>
- [3] Hegel, G.W.F.: *Encyklopädie der philosophischen Wissenschaften im Grun-drisse*. Heidelberg, 1817. Rev. Ed. Berlin, 1827, 1830.
- [4] Kelemen, J.: *Postmoderný stroj*. FR&G, Bratislava 1998

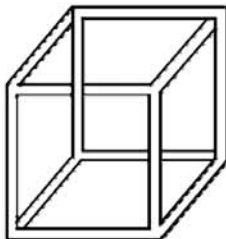
Paradoxy vizuálnej percepcie

Ján Rybár¹

Abstrakt. Článok sa zaoberá skúmaním paradoxných vizuálnych objektov, ktoré súvisia najmä s heuristikami vnímania hĺbky. Súčasťou týchto heuristik je celá paleta indikátorov monokulárneho vnímania 3D. Skúmanie týchto indikátorov je cesta, ktorá vedie k vysvetleniu vizuálnych paradoxov.

Úvod

Na rôznych webových stránkach, na plagátoch, kde ide o upútanie pozornosti diváka sa často môžeme stretnúť s rôznymi variantmi tzv. nemožnej kocky (obr. 1.) V každom prípade je to zvláštna kocka.² Ak si budeme všímať len jednotlivé detaily obrázku, bude všetko v poriadku, ak sa však pozrieme na kocku ako celok, je to nezmysel. Jeden z jej zadných „pilierov“ je postavený pred predným „nosníkom“. Základnou vlastnosťou paradoxnej kocky je, že hoci sa dá sa nakresliť v dvojdimenzionálnom priestore (2D), nedá sa skonštruovať v trojdimenzionálnom priestore (3D). Podobných obrazcov je skonštruovaných mnoho, niektorými sa budeme ďalej zaoberať podrobnejšie. Ide vlastne o špeciálny druh optických ilúzií, ktoré nie náhodou dostali meno nemožné objekty.



Obrázok 1. Nemožná kocka.

Je zaujímavé, že skúmanie nemožných objektov môže efektívnym spôsobom prispieť k objasneniu vizuálnych paradoxov obzvlášť tých, ktoré súvisia s vnímaním hĺbky. Toto skúmanie môže pomôcť odpovedať na mnohé základné otázky: Ako je možné, že dokážeme vnímať

¹ Centrum kognitívnych vied, Katedra aplikovanej informatiky FMFI UK, Bratislava, e-mail: rybar@fmph.uniba.sk

² V podstate ide o upravenú Neckerovu kocku. Prvú známu verziu nemožnej kocky prezentoval holandský grafik M. C. Escher (1898-1972) vo svojej známej práci *Belvédér*. Autor tu prejavil značnú dávku zmyslu pre humor, pod belvédérom sedí na lavičke chlapec, pred sebou má nákres nemožnej kocky, a zároveň ju už zostrojenú drží v rukách. Najjednoduchšiu verziu nemožnej kocky, ktorú uvádzame v tomto článku, možno nájsť na stránke: <http://mathworld.wolfram.com/FreeMishCrate.html>

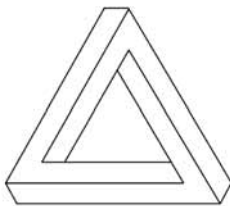
priestor, hoci na sietnici máme v podstate len 2D zobrazenie objektov? Prečo náš vizuálny systém má tendenciu niektoré dvojrozmerné objekty vnímať ako trojrozmerné? Ako je možné, že na 2D (teda na ploche) vieme vytvoriť ilúziu 3D?

Je známe, že dávno pred vznikom vedeckej psychológie vizuálne paradoxy boli predmetom pozornosti výtvarných umelcov. A ako uvidíme ďalej v súvislosti s konštrukciou nemožných objektov mnohých výtvarníkov, tieto vizuálne paradoxy fascinujú dodnes.

Typy paradoxných vizuálnych objektov

Začiatky konštrukcie paradoxných „nemožných“ objektov, resp. špeciálnych priestorových efektov na ploche môžeme nájsť už v renesančnom umení (Pieter Breughel a jeho obraz *Zlodej na šibenicí*, r.1568 je toho dobrým príkladom).³

Na pôdu vedeckej psychológie sa problematika dostala až v päťdesiatych rokoch minulého storočia, keď Lionel S. Penrose a jeho syn Roger Penrose publikovali článok v *British Journal of Psychology*, v ktorom boli nakreslené a popísané dva typy nemožných objektov. Prvý typ objektov, ktorý popisujú je tzv. nemožný (paradoxný) trojuholník. Každý uhol tohto trojuholníka sa javí ako perfektne pravý, ale celkove je to veľmi zvláštny trojuholník (pozri obr. 2).



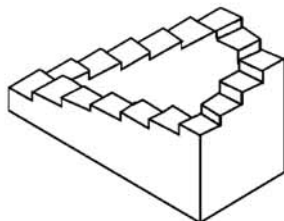
Obrázok 2. Penrosov nemožný trojuholník.

Nezávisle od Penrosovcov skonštruoval podobný trojuholník aj švédsky výtvarný teoretik Oscar Reuterswärd (1915-2002). Prvý nemožný trojuholník nakreslil z kociek už v roku 1934. Je autorom viac než dvetisíc rôznych variácií tohto typu nemožných objektov. Dokonca

³ Na Breughelovom obraze vrchná časť šibenicí je v inej perspektíve ako spodná. O niečo podobného, len v trochu komplikovanejšej podobe, sa pokúsil Escher v spomínanej grafike *Belvédér*. Poschodia letohrádku (vyhliadkovej veže) sú v rôznej perspektíve. Je to urobené veľmi rafinovane. Na prvom poschodí niektoré piliere majú spodnú časť vpredu a hornú časť vzadu (pozri Escherovu oficiálnu web-stránku na adrese: <http://www.mcescher.com/Gallery/gallery.htm>). I keď v tomto prípade ide o paradoxný objekt, nejde o nemožný objekt v pravom zmysle slova, pretože takýto objekt sa dá reálne postaviť.

vo Švédsku bola vydaná na jeho počesť poštová známka s námetom nemožného trojuholníka.

Druhým slávnym nemožným objektom, o ktorom sa hovorí v článku Penrosovcov sú tzv. večné schody (pozri obr. 3). Po týchto schodoch môžeme neustále vystupovať alebo zostupovať.⁴ Je to jeden z najznámejších nemožných objektov. Ilúzia sa podľa mien jej autorov často nazýva aj Penrosove schody.



Obrázok 3. Penrosove schody.

Z ďalších nemožných objektov je populárna tzv. diabľova vidlička, respektíve trojzubec (obr. 4). Ak sa pozrieme len na ľavú alebo len na pravú stranu vidličky, všetko bude v poriadku, avšak ako celok je to typický nemožný objekt. Základným princípom konštrukcie tohto nemožného objektu je, že na znázornenie valca potrebujeme dve čiary a kruh, ale na znázornenie hranola potrebujeme tri čiary (ako to vidíme na obr. 4.).⁵



Obrázok 4. Diabľova vidlička.

Tento paradox má mnoho variantov. Medzi najznámejšie patrí rímsky vojak, ktorý stojí pod víťazným oblúkom, pričom nevieme presne rozhodnúť či má oblúk dva alebo tri piliere. Podobne je to pri Shepardovej nemožnej figúre slona, o ktorom nevieme povedať koľko má presne nôh (pozri Shepard 1990).

M. C. Escher a vizuálne paradoxy

Je zaujímavé, že Eschera mnohí jeho súčasníci považovali skôr za dobrého remeselníka, než výtvarného umelca. Dokonca boli vy-

⁴ Profesor Zimbardo vo svojom televíznom kurze psychológie uvádza dokonca animovanú podobu tejto ilúzie (po schodoch smerom hore alebo dole večne skáče guľička).

⁵ Pozri: Seckel 2006. Obrázok rozdelenej vidličky je prevzatý z práce K. Rebrovej.

slovené pochybnosti či vôbec ide o ešte o výtvarné umenie, alebo sú to len rôzne intelektuálne hračky a vedecké záhady. Escher dosiahol väčšiu popularitu medzi prírodovedcami a matematikmi než medzi výtvarníkmi a výtvarnými teoretikmi.⁶ Nie je to náhodné, keďže je o ňom známe, že svoje inšpirácie hľadal (hoci nemal žiadne formálne prírodovedné vzdelanie) práve v tejto oblasti. Veľký vplyv mal na neho jeho brat, ktorý bol kryštalograf. Escher svojim skúmaním a kreslením mnohostenov vzbudil záujem aj u profesionálnych prírodovedcov, ktorí ho pozývali prednášať o symetrii na svoje vedecké semináre. Ovplyvnilo ho aj čítanie niektorých matematických odborných článkov, medzi nimi predovšetkým Polyov článok o symetrických grupách a takisto článok profesora Coxetera, ktorý mal významný vplyv na tvorbu jeho grafík súvisiacich s princípom nekonečna. V každom prípade Escher mal výnimočnú schopnosť vizualizovať také matematické princípy, ako sú symetria, rekurzia, limita a ďalšie. Dokázal sa zmocniť aj takej témy ako je princíp autoreferencie. Ukážkovou toho je litografia s názvom *Galéria*.⁷

Escher po štúdiách na Haarlemskej škole architektúry a dekoratívneho umenia odišiel do Talianska, kde žil a tvoril až do roku 1935, keď klíma Mussoliniho režimu sa stala pre neho už neznesiteľná a po dočasnom pôsobení na rôznych miestach sa nakoniec usadil vo svojom rodnom Holandsku. Prestáhanie na sever znášal Escher spočiatku veľmi ťažko. V Taliansku ho fascinovali stredomorské scenérie a kultúra a jeho hlavnou témou boli krajinky. Keď sa musel presťahovať zistil, že krajina na severe je pre neho fádna. To bol pravdepodobne jeden z dôležitých impulzov pre hľadanie nových tém, pre posunutie sa k tomu o čom jeho životopisci píšu ako o „mental imagery“. A až v tomto období vznikajú jeho najslávnejšie práce venované vizuálnym paradoxom.

V tejto súvislosti treba uviesť, že R. Penrose mu poslal svoj článok o nemožných objektoch, pretože vedel o jeho záujme o tieto paradoxy z Escherovej výstavy, usporiadanej pri príležitosti svetového matematického kongresu v Amsterdame. A urobil dobre, pretože Escher sa skutočne dokázal inšpirovať a zmocniť témy nemožných objektov veľmi originálnym spôsobom.

Penrosove paradoxné schody sa stali námetom jeho slávnej grafiky *Ascendentné a descendentné* (niekedy nazývanej aj *Mnísi*). Gra-

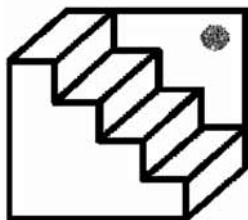
⁶ V tejto súvislosti nezabudnime pripomenúť, že je stále veľmi populárny aj medzi laickou verejnosťou. Mnohé jeho práce „zľudoveli“ a často jeho obdivovatelia ani nevedia, že ich autorom je Escher.

⁷ Pozri: <http://www.mcescher.com/Gallery/gallery.htm>

fika predstavuje mníchov na večnom schodisku. Po jednej strane schodiska mnísi stále zostupujú a po druhej strane toho istého schodiska iní mnísi stále vystupujú v uzatvorenom cykle.

Veľkým námetom sa stal pre neho aj nemožný trojuholník. Túto ilúziu spracoval v slávnej grafike *Vodopád*.⁸ Táto grafika je skvelým príkladom, že jeho práce sú viacvrstvové. Ak si túto grafiku pozorne prezrieme, tak zistíme, že okrem efektu vody, ktorá paradoxne tečie hore kopcom, grafika obsahuje tri nemožné trojuholníky. Pričom sú konštruované tak, že ak si pozrieme len ľavú alebo len pravú stranu práce, tieto trojuholníky neuvidíme. Vidíme ich len, keď pozeráme na prácu ako celok.

Escher bol ozajstným majstrom vizuálnych paradoxov. Jeho obľúbenou technológiou bolo premyslené využívanie optických ilúzií, napríklad dvojznačných figúr. V tomto smere vynikajúco uplatnil dvojznačnú optickú figúru *Schroederove schody* (pozri obr. 5). Ak sa pozornejšie pozrieme na dvojznačný obrazec po chvíli zistíme, že tá strana schodiska, na ktorej je okrúhly znak môže byť zadnou stenou alebo prednou stenou schodiska (v druhom prípade sú schody otočené dolu hlavou). Ak si vopred na tejto figúre nacvičíme preklikanie z jedného obrazca na druhý, budeme pripravení pochopiť, prečo je na jeho grafike *Relativita* možné, že dve postavy idú jedným smerom na jednom schodisku, a predsa jedna ide dole a jedna hore. Tá istá dvojznačná figúra je viacnásobne uplatnená aj v grafike *Konvexné a konkávne*, a potom sa naozaj tie isté schody zdajú raz vypuklé a raz preliačené. V Escherových grafikách môžeme nájsť také paradoxné objekty, akým je rebrík, ktorý je zároveň vonku aj vo vnútri (pozri grafiku *Belvédér*) alebo podlahu, ktorá môže byť súčasne aj stropom (pozri grafiku *Podlubie*) a ďalšie.⁹



Obrázok 5. Schroederove schody.

⁸ Obidve spomínané grafiky môžeme nájsť na stránke: <http://www.mcescher.com/Gallery/gallery.htm>

⁹ Aj tieto uvedené grafiky môžeme nájsť na stránke: <http://www.mcescher.com/Gallery/gallery.htm>. Tiež ich môžeme nájsť v reprezentačných monografiách venovaných Escherovmu grafickému dielu (Locker 1992, Locker 2000).

Avšak najdôležitejšie u Eschera sú paradoxy spojené s vnímaním hĺbky (3D). V žiadnom prípade nejde o triviálny problém. Ako je známe, na sietnici sú objekty zobrazené v podstate dvojrozmerné a napriek tomu ich vidíme trojrozmerné. Ako je to možné? Zvyčajná odpoveď je táto: „Pretože sme binokulárni“. Samozrejme, odpoveď ide správnym smerom. Ak sa objekt približuje k nám, oči sa otáčajú k nosu. Toto otáčanie (binokulárna konvergencia), pohyb očných svalov je významným indikátorom pre mozog pri odhadovaní vzdialenosti objektu. Druhým takým indikátorom je fakt, že vnímaný obraz z oboch očí nie je celkom totožný (binokulárna disparita). Môžeme sa o tom ľahko presvedčiť, keď sa pozrieme na objekt, ktorý je blízko pred nami najprv jedným a potom druhým okom.¹⁰ Na základe rozdielných informácií z jedného a druhého oka mozog indikuje vzdialenosť objektu v priestore. Konvergencia a disparita sú nepochybne dôležitými faktormi priestorového vnímania.

To však ešte nemôže byť úplným vysvetlením vnímania hĺbky, a to z veľmi jednoduchého dôvodu – hĺbku dokážeme vnímať aj jedným okom.

Paradoxy a monokulárne vnímanie hĺbky

V tejto súvislosti nás zaujímajú vlastne dve otázky: 1. v čom spočíva podstata vnímania hĺbky jedným okom. 2. Ako sa konštruuje hĺbka (ilúzia hĺbky) v dvojrozmernej rovine. Na základe čoho teda môžeme vytvárať a vnímať trojrozmerné efekty v rovine. Odpoveď na prvú i druhú otázku treba hľadať v manipulácii s monokulárnymi indikátormi vnímania hĺbky.

Escher priam inžiniersky presne pracuje s týmito indikátormi, dokáže pomocou nich rafinovane popliesť náš perцепčný systém. Paradoxy, s ktorými pracuje obyčajne nie sú prvoplánové, až keď si všimneme obrázok bližšie (a obrázok vždy vyžaduje bližšie skúmanie) vidíme, že niečo nie je v poriadku. Čím detailnejšie ho budeme skúmať tým hlbšie sa budeme dostávať do siete jeho kognitívnych hier.

Znamenitým príkladom je už spomínaná Escherova grafika *Vo-dopád*.¹¹ Na grafike je prezentovaný mlyn. Ak sa budeme na obrázok pozerat' analyticky (kúsok po kúsok pôjdeme vodným kanálom) všetko bude v poriadku – voda tečie stále dole. Ak sa pozrieme na obrázok ako celok, budeme vidieť niečo nezmyselné a nemožné, že voda tečie hore kopcom. Na obrázku nás šokuje konflikt medzi lokálnym a globálnym prístupom.

¹⁰ Tu treba poznamenať, že ak je objekt vzdialený od nás viac než sto metrov, rozdiely sú už zanedbateľné a v tom zmysle sme vlastne monokulárni.

¹¹ Pozri: <http://www.mcescher.com/Gallery/gallery.htm>

Na prvý pohľad vyzerajú Escherove nemožné objekty ako veľká fantázia. Keď sa však na ne pozrieme cez psychológiu vnímania zistíme, že sú zostrojené podľa prísnych pravidiel, že rešpektujú pravidlá (presnejšie indikátory) zobrazovania hĺbky. Prvým takým indikátorom je tu prekryvanie objektov. Hoci len čiastočne prekryté objekty sa na dvojrozmernom obraze vždy zdajú vzdialenejšie. Druhý indikátor súvisí s umiestnením objektov vo výške. Objekty umiestnené vyššie (nad horizontom) sa zdajú vzdialenejšie, objekty umiestnené nižšie (pod horizontom) sa zdajú bližšie. V danom prípade je využitý výškový efekt, a napriek tomu, že kanál smeruje v rovine do výšky, máme dojem, že voda tečie do hĺbky. Tretí indikátor, ktorý pomáha vytvárať hĺbku sú konvergujúce línie, teda lineárna perspektíva. Je to najdôležitejší indikátor, s ktorým Escher v grafike pracuje. Štvrtým indikátorom je veľkosť objektov na obraze. Väčšie objekty na grafike sa zdajú bližšie než menšie (ten istý mnohosten na pravej veži mlyna je väčší než mnohosten na ľavej strane, aj preto sa prvý zdá bližšie než druhý). Prúd vody v kanále tečúci do hĺbky je posilnený zmenšovaním veľkosti múrika, ktorý vodu obklopuje. Dôležitým indikátorom je povrchová štruktúra nakresleného múrika kanála aj samotnej vody. Je hrubšia, zrnitejšia na začiatku kanála a jemnejšia na konci, čo takisto posilňuje ilúziu hĺbky. Ďalším indikátorom je tienenie. Tmavšie časti objektu (napríklad spomínané múriky vodného kanála) sa zdajú byť vzdialenejšie než svetlejšie oblasti. Posledným indikátorom je vzdušná perspektíva. Obrazy objektov, ktoré sú nejasnejšie sa zdajú vzdialenejšie, než objekty s výraznými zreteľnými obrysami. Escher práve pomocou majstrovského vyžívania indikátorov monokulárneho videnia hĺbky dokáže vytvoriť ilúziu, že voda tečie hore kopcom. Na začiatku sa nám mohlo zdať, že ide o záhadný výtvar autorovej výstrednej mysle. Teraz vidíme, že to nie je tak. Escherov prístup je vysoko racionálny.

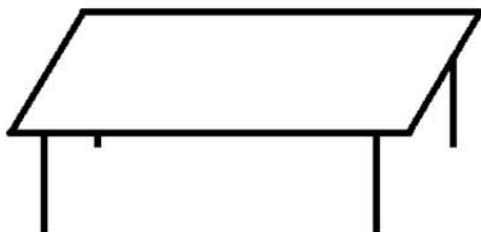
Paradoxy a perspektíva

Ako sme už spomínali kľúčovým indikátorom vnímania hĺbky je lineárna perspektíva. Ako poznamenáva R. Gregory (2004), trvalo veľmi dlho kým sa človek prepracoval k objavu lineárnej perspektívy. Tento autor radí objav perspektívy medzi také prevratné objavy v dejinách civilizácie akými boli objavy ohňa alebo kolesa. Ako je známe, k objaveniu perspektívy došlo až v renesančnom období. Staré egyptské umenie, antika, stredovek nepoznajú perspektívu.¹² Ume-

¹² To, že sa s lineárnou perspektívou začalo pracovať tak neskoro, je o to paradoxnejšie, keď si uvedomíme, že na sietnici zobrazenie objektov je v perspektíve. Každým zdvojením vzdialenosti objektu od oka, je zobrazenie veľkosti objektu na sietnici o polovicu menšie (takzvaný Emmertov zákon).

lecký kanón tých čias napríklad predpisoval zobrazovať veľkosť objektov podľa dôležitosti. Položme si hypotetickú Gregoryho otázku: Čo by sa stalo keby perspektíva bola objavená neskôr než fotografia? Odpoveď je jednoduchá: fotografia by sa určite nepovažovala za verný obraz skutočnosti.

Rafinovaná práca s lineárnou perspektívou je základom pri konštrukcii mnohých paradoxných optických ilúzií. Týka sa predovšetkým tých ilúzií, ktoré vznikajú porušením konštantnosti tvaru a veľkosti. Tieto efekty dosiahneme, ak dva rovnaké objekty umiestnime za sebou na pozadie v lineárnej perspektíve (na nejakých líniách, ktoré konvergujú do jedného bodu). Napriek tomu, že objekty sú rovnaké, podľa umiestnenia v lineárnej perspektíve sa nám môžu zdať väčšie alebo menšie. Napríklad Shepardove monštrá, ktoré sú rovnaké, ale nezdájú sa takými vďaka tomu ako sú umiestnené na pozadí v lineárnej perspektíve. Lineárna perspektíva je základným ilúziotvorným (paradoxotvorným) faktorom aj v ďalšej slávnej Shepardovej ilúzii, v takzvaných otočených stoloch. Ide o rovnaké stoly, z ktorých jeden je otočený viac k nám po dĺžke a druhý je otočený viac po výške. Efekt nerovnakosti spôsobuje to, že vnútorné hrany stolov sú v perspektíve, čo v jednom prípade ešte predlžuje dlhšiu a v druhom prípade zase rozširuje širšiu stranu stola (Shepard 1990, Ninio 2001).¹³



Obrázok 6. Efekt neuplatnenia perspektívy.

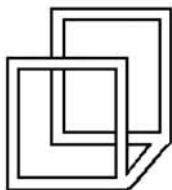
K zaujímavým výsledkom však dochádza aj pri ignorovaní zobrazovania objektov pomocou lineárnej perspektívy. Ak napríklad plocha stola v obdĺžnikovom tvare nie je nakreslená v perspektíve, horná hrana sa bude zdať dlhšia než spodná a bočné hrany už nie sú celkom rovnobežné, pri pozornejšom pohľade sa nám bude zdať, že divergujú (pozri obr. 6).¹⁴

¹³ K ilúzii nerovnakosti stolov prispieva do určitej miery aj fakt, že vertikály sa nám subjektívne zdajú dlhšie než horizontály.

¹⁴ Pozri: Gregory 2004 a Ninio 2001. Efekt je ešte zosilnený hrou s nerovnakým prekrytím zadných nôh stola a tiež, že nohy stola delia spodnú hranu stola na viac častí, čo ju tiež skracuje.

Vrodené alebo naučené?

Prečo útvar na obr. 7 vidíme ako priestorový objekt a nie iba ako niekoľko štvorcov prepojených dvoma čiarami? Vnímanie priestorových útvarov v rovine viedlo k otázke či ide o schopnosti vrodené alebo naučené.



Obrázok 7. Demonštrovaný obrázok.

V tejto súvislosti uvádza zaujímavý výskum vo svojom článku E. Gál (1994).¹⁵ V experimente boli utvorené dve skupiny pokusných osôb. Jednu tvorili Európania, druhú príslušníci afrického kmeňa Zuluov. Obidvom skupinám bol demonštrovaný obrázok 7. Úlohou pre obidve skupiny bolo postaviť tento útvar zo zápaličiek. Európania ho urobili priestorovo a Zuluovia na ležato. Zdá sa, že aj tieto výsledky potvrdzujú vnímanie priestoru ako do značnej miery kultúrne podmienené.

Vedecká komunita však ani zďaleka nemá na túto otázku jednotný názor. Svojho času vznikla o tejto téme zaujímavá diskusia medzi J. Fodorom, tvorcom teórie modularity, podľa ktorej základnou vlastnosťou nízko-úrovňovej kognície je informačná uzatvorenosť a P. Churchlandom, stúpencom teórie *ladenness*, podľa ktorej je vnímanie ovplyvnené našou predchádzajúcou skúsenosťou a teoretickým backgroundom (Fodor 1984, Churchland 1988).

J. Fodor ukazuje svoje argumenty na tak banálnom fenoméne akým je Müller-Lyerová figúra (ide o rovnaké úsečky, avšak jedna úsečka je zakončená krídelkami dovnútra a druhá navonok, a preto úsečka s krídelkami navonok sa zdá dlhšia). Pri pohľade na túto figúru vidíme ešte aj ďalšiu ilúziu. Úsečku s krídelkami navonok vidíme ako vnútorný roh a úsečku s krídelkami dovnútra ako vonkajší roh nejakej miestnosti. Ilúzie vzdorujú, aj keď vieme, že sú to ilúzie, a tento poznatok, táto vedomosť ich neodstraňuje. Na základe toho Fodor usudzuje, že percepcia (prinajmenšom jej časť), nie je penetrovaná kogníciou a je informačne uzatvorená. Fodor z toho odvodzuje aj ďalšie dôležité epistemologické dôsledky. Fixnosť a neutrálnosť percepcie vo vzťahu k centrálnym kognitívnym procesom je základ-

¹⁵ Pozri tiež Monaco, J.: Jazyk filmu: Znaky a syntax. In: *Sbornik filmové teórie I*, Praha 1991.

nou platformou pri jeho kritike teórie *ladenness* a epistemologického relativizmu, ktorý je spätý s touto teóriou.

Fodorova platforma má však aj tvrdých kritikov, napríklad spomínaného Paula Churchlanda. Podľa neho Müller-Lyerova figúra a iné druhy optických ilúzií sú kultúrne nadobudnuté zručnosti (ako o tom svedčí aj uvedený výskum s príslušníkmi kmeňa Zulu).

Napriek tomu Fodor [2] odmieta názor, že optické ilúzie sú kultúrne determinované, ako aj názor, že by penetrácia percepcie centrálnymi kognitívnymi procesmi mohla byť záležitosťou tréningu a učenia, a to stále z toho istého dôvodu – percepčné moduly sú informačne uzatvorené.

Nepridať sa ani na jednu, ani na druhú stranu a na záver tvrdiť, že problém je v tom, že obidve strany hyperbolizujú svoj prístup a pravda je niekde uprostred, by bolo síce možno správne, ale trochu lacné. Zaujímavé na tom je, že aj Fodor [2] k svojmu pohľadu prichádza len vďaka tomu, že jeho teoretické pozadie je iné, že jeho teória *ladenness* je iná a to mu umožňuje aj „vidieť“ inak.

Záver

Primárne vysvetlenie vizuálnych paradoxov treba hľadať v tom, že percepčný systém „neskenuje“ a nespracováva všetky informácie zo zorného poľa. Prakticky to ani nie je možné, pretože systém by sa veľmi rýchlo „zahltil“ (veľmi rýchlo by sa vyčerpala jeho kapacita). Percepčný systém spracováva informácie len výberovo a táto skutočnosť ho potom „núti“ používať určité odhadovacie heuristiky. Uplatnenie týchto heuristik vedie k vzniku najrôznejších vizuálnych paradoxov. Našťastie tieto heuristiky (resp. indikátory) majú pomerne stabilnú štruktúru a preto sú predmetom rozsiahleho výskumu.

V každom prípade odhadovacie heuristiky (a v tom zmysle aj vizuálne paradoxy) svedčia nie o nedostatkoch, ale skôr o prednostiach a inteligencii percepčného systému.

PodĎakovanie: Tento príspevok vznikol v Centre kognitívnych vied, KAI, FMFI UK v Bratislave a bol podporený VEGA MŠ SR, projekt č. 1/3612/06.

Literatúra

- [1] Churchland, P. (1988): Perceptual Plasticity and Theoretical Neutrality: A Reply to Jerry Fodor. *Philosophy of Science*, **55**, 1988, s. 176.
- [2] Fodor, J. (1984): Observation Reconsidered. *Philosophy of Science*, **51**, s. 34.
- [3] Gál, E. (1994): Mechanizmy klasifikácie. In: *Kapitoly z epistemológie II*. Bratislava: Univerzita Komenského.
- [4] Gregory, R. L. (2004): *Eye and Brain. The Psychology of Seeing*. Fifth edition. Oxford University Press.
- [5] Locker, J. L. (Ed.) (2000): *The Magic of Escher*. New York: Harry N. Abrams, 2000.

- [6] Locker, J. L. (Ed.) (1992): *M. C. Escher: His Life and Complete Graphic Work*. New York: Harry N. Abrams.
- [7] Ninio, J. (2001): *The Science of Illusions*. Cornell University Press.
- [8] Penrose, L. S., Penrose, R. (1958): Impossible Objects: A special type of visual illusion. *Brit. J. Psychology* **49**, 31-33.
- [9] Seckel, A. (2006): *Optical illusions. The Science of Visual Perception*. New York: Firefly Books.
- [10] Shepard, R. (1990): *Mindsights. Original visual illusions, ambiguities, and other anomalies with a commentary on the play of mind in perception and art*. New York: W. H. Freeman and company.

Modely usudzovania

Ján Šefránek¹

Abstrakt. Načrtnutý je pojmový rámec, v ktorom podstatnou súčasťou mysle je usudzovanie a modely² usudzovania sú jadrom modelov mysle. Pozornosť sa sústreďuje na rôzne prístupy k štúdiu (modelovaniu) usudzovania. Prezentujú sa modely usudzovania, ktoré možno konštruovať s využitím prostriedkov logiky (matematickej, neklasických, nemonotónnych), pozornosť sa venuje otázke, nakoľko je logika o usudzovaní. Ďalej sa pozornosť prepína na alternatívne modely usudzovania, ktorých ambíciou je pracovať aj s inými reprezentáciami (grafickou, heterogénnou, rozlišovacími kritériami, ktoré možno predpokladať aj v živočíšnej ríši) a so zodpovedajúcimi spôsobmi usudzovania.

1 Úvod

Najprv si vymedzme hraciu plochu (základné predstavy). Predmetom záujmu budú *kognície* (budem ich nazývať aj *agentami*). Čo pod kogníciou rozumiem, budem postupne skladať v tomto úvode. Každá kognícia pôsobí v nejakom *prostredí*. Súčasťou prostredia je aj organizmus daného agenta. O *organizme* sa bude predpokladať iba toľko, že sa môže nachádzať v rôznych stavoch. Táto predstava pripúšťa aj netelesného agenta - napríklad program s množinou jeho možných stavov.

Kognícia (agent) je v priamom kontakte s prostredím cez senzory a aktuátory. Pomocou *senzorov* zbiera informácie o tej časti prostredia, ktorá je v ich (limitovanom) dosahu v danom čase. Podotýkam, že senzory môžu zbierať informácie o stave vlastného organizmu. Kognície zasahujú do prostredia pomocou *aktuátorov* (menia prostredie alebo menia svoju polohu v ňom).

Dôležitým komponentom navrhovanej predstavy o kognícii je *energia a reprezentácia*.

K pojmu energie najprv príklad. Agent (kognícia) A chce zistiť, či pri montáži keramického obkladu vane sa myslelo na potrebu prístupu k sifónu a odtokovému kanalizačnému potrubiu v prípade poruchy. Agent B poradí: “poklep, či niektorá obkladačka neznie duto”. Agent B začne preklepávať od stredu. Agent A: “skús pri odtoku”. Agent B okamžite: “no jasné, ja som blbec”.

Príklad ukazuje, že niektoré kognície sú živšie, iné pasívnejšie, lenivejšie. Niektoré z tých lenivejších, ak dostanú impulz, dokážu usu-

¹ Fakulta matematiky a informaiky, Univerzita Komenského, e-mail: sefranek@ii.fmph.uniba.sk.

² Slovo model sa v texte používa v dvoch významoch. Jednak v technickom – tak, ako sa používa v logickej sémantike. Jednak v neformálnom – ako teória, poňatie nejakej domény, abstrakcia takejto domény. Verím, že čitateľ obe použitia ľahko rozlíši.

dzovať približne tak ako tie živšie. Samozrejme, táto “živost” alebo “lenivosť” závisí na kontexte (čase, objekte záujmu, stave organizmu a podobne). Spomínaná živost alebo pasívnosť je jedným z prejavov *energie*. Ďalšími dôležitými (a súvisiacimi) sú emócie (emotívne stimuly pre aktiváciu a zacielenie kognície), motivácia, vytrvalosť, schopnosť zamerať pozornosť, schopnosť koncentrovať sa, schopnosť nájsť vhodný uhol pohľadu atď. atď. V tomto texte nebudem venovať pozornosť energii. Neznamená to však, že si neuvedomujem jej dôležitosť, dokonca si myslím, že konštrukt energie môže mať miesto aj v modeloch usudzovania. Na dôležité aspekty kognitívnych schopností, ktoré tu pokrývam konštruktom energie, upozorňuje vo svojich originálnych prácach Ladislav Kováč, pozri napríklad (2006, 2004)

Prejdime ku konštruktu *reprezentácie*. Podstatnou schopnosťou organizmov, ktoré niečo *vedia*, je schopnosť *rozlišovať* (čo je jedlé, čo je nebezpečné, kde môže byť nejaká potrava, v akej situácii použiť nejakú frázu, aký argument je korektný, čo je riešenie danej diferenciálnej rovnice atď.).

Je rozumné predpokladať: ak agent (kognícia) niečo vie, opiera sa o nejakú (mentálnu)³ štruktúru, v ktorej má tú znalosť zakódovanú. V kontraste k tomu, ak agent niečo náhodne dosiahol, zaobídeme sa bez takéhoto predpokladu. Ak agent čosi nevie, je neprimerané predpokladať, že chýbajúcu znalosť má nejakou zakódovanú. Napríklad: ak robot dokáže (vie) zbierať prázdne plechovky od coca-coly, ale nie tie od piva a iba raz sa prevrátil, je zrejme, čo doňho konštruktéri zakódovali a čo nie.

Schopnosť rozlišovať (vedenie, poznanie, rozumenie) preto môžeme vysvetliť konštruktom reprezentácie. *Reprezentácia* je tá zakódovaná štruktúra, o ktorú sa agent opiera, keď niečo vie. Ak agent disponuje reprezentáciou toho, čo je jedlé, potom dokáže rozlišovať jedlé veci od nejedlých; ak agent nedisponuje reprezentáciou toho, čo je nebezpečné, nedokáže rozlišovať nebezpečné udalosti atď. Keď máme pojem, s použitím ktorého dokážeme vysvetliť, prečo nejaký agent niečo vie a iný to nevie, naša schopnosť opisu kognícií sa zosilňuje. Pojem reprezentácie používame práve preto, aby sme dokázali kognície opisovať detailnejšie.

V tomto texte sa používajú rôzne úrovne opisu reprezentácií. Za základnú však považujem úroveň rozlišovacích kritérií, abstrakcií schopnosti rozlišovať, podrobnejšie pozri v (Sefránek 2002) alebo (Takáč 2006a-c.2007a,b). Nad touto úrovňou dokážeme nadstavovať ďalšie.

³ Bez predpokladu, že ide o mentálnu štruktúru sa pohodlne zaobídeme. Argumentáciu pozri v (Sefránek 2002).

Teraz je čas na podstatný krok - odlišiť reaktívnu kogníciu a myseľ.

Reprezentácia, ktorou disponujú *reaktívne kognície*, iba sprostredkuje prenos informácie medzi senzormi a aktuátormi. Niektoré (nie nutne všetky) signály, registrované senzormi, aktivujú nejaké rozlišovacie kritériá. Aktuátory sa aktivujú (alebo neaktivujú) v závislosti na výstupoch rozlišovacích kritérií (opäť zdôrazňujem, že, niektoré rozlišovacie kritériá zaznamenávajú vnútorný stav kognície). Podstatné je, že reprezentácia (rozlišovacie kritériá) umožňuje prepojiť stav prostredia, v ktorom sa vyskytuje agent, sprostredkovaný senzormi s konaním agenta. Gärdenfors (1995) charakterizuje takúto reprezentáciu prívlastkom *cued*. Agent prostredníctvom takejto reprezentácie iba reaguje na momentálny stav tej časti prostredia, ktorú snímajú jeho senzory a na ktorý (stav) sú senzitívne jeho rozlišovacie kritériá. O možnostiach modelovania na báze reaktívnych agentov pozri (Lúčny 2007).

Ideový konštrukt reprezentácie v tomto prípade potrebujeme iba na to, aby sme opísali prepojenie senzorov a aktuátorov. Bez takéhoto (alebo podobného) konštraktu by sme mali problémy rozlíšiť, na aké podnety sú kognície senzitívne do tej miery, že reagujú na ne akciami a vysvetliť túto schopnosť (ako kognitívnu schopnosť).

Kogníciu budeme považovať za *mysleľ* vtedy, keď disponuje rozlišovacími kritériami, ktoré reagujú na stav (jej) iných rozlišovacích kritérií. Inými slovami, ak sa v reprezentácii spracúva aj momentálny stav vlastnej reprezentácie. Pre myslenie (ako proces a schopnosť), ale aj myseľ (ako stav a produkt) je charakteristické práve spracovávanie vlastnej reprezentácie, operácie na reprezentáciách. Operácie na reprezentáciách budeme nazývať aj *usudzovaním*. Dostali sme sa k hlavnej téme nášho textu a bude nás zaujímať akými rozmanitými spôsobmi možno usudzovanie opisovať (modelovať, charakterizovať).

V ďalšom sa najprv budeme baviť o logike ako modeli usudzovania. Kým sa dostaneme k detailom, jednu zásadnú poznámku: Nedomnievam sa, že pri analýze a opise reprezentácie a usudzovania je potrebné prijímať predpoklady a záväzky ontologickej povahy. Ak predpokladáme reprezentáciu pomocou rozlišovacích kritérií, príliš nás nemusí zaujímať, či nejaké rozlišovacie kritéria skutočne existujú v kognitívnom systéme (napr. človeka). Podobne si nemyslím, že je potrebné predpokladať existenciu nejakých syntakticky štrukturovaných logických reprezentácií v kognitívnom systéme.

Model, charakterizujúci kogníciu ako výpočty nad syntakticky štrukturovanými reprezentáciami, ilustruje Fodor (2000) takto: “The main idea of rationalist psychology is that beliefs, desires, thoughts,

and the like have logical forms, and that their logical forms are among the determinants of the roles they play in mental processes. For example *John swims and Mary drinks* is a conjunctive belief, and that is why having it can lead one to infer that *John swims ...*". Predstava o existencii konjunktívnych elementov v reprezentácii (takto treba kvalifikovať *conjunctive beliefs*) je minimálne v rozpore s Ockhamovou britvou. Ak máme v reprezentácii uvedený konjunktívny element, mali by sme mať aj niečo také ako *John swims and Jack sleeps and Mary drinks*, ale aj *Jack sleeps and Mary drinks* a tak podobne do nekonečna. Okrem toho a hlavne, tvrdenie o tom, ako z uvedenej konjunkcie (v reprezentácii) odvodzujeme konjukty (mentálnymi procesmi na základe logickej formy) sa prieči introspektívnej evidencii. Aby sme prešli od introspekcie k náčrtu spoľahlivejších metód: Možno si predstaviť experimenty, pri ktorých experimentálne subjekty do protokolu zaznamenávajú detailne priehy svojich myšlienok. Okrem toho, máme k dispozícii bohatý materiál – publikované polemiky, protokoly zo súdnych pojednávanií a pod. Neverím, že by sme v nich našli odvodenia konjunktov z vyslovených konjunktív. Neverím ani, že by sa podobné kroky odvodenia vyskytli v dôkazoch "pracujúceho matematika". Zhrniem a zopakujem: logický model (a žiadny iný model) usudzovania netreba považovať za opis s ontologickou záväznosťou. Stačí, keď nám slúži ako užitočný nástroj. V texte budem pri jednotlivých modeloch formulovať predstavy o ich funkciách. Na tomto mieste iba prvý komentár k účelu modelu rozlišovacích kritérií. Dajú sa použiť na vyjadrenie hypotéz o evolúcii kognície v prírode, u individua, u kooperujúcich individuí a dajú sa nad nimi nadstavovať iné modely. V neposlednom rade, tento model bol využitý vo viacerých zaujímavých simuláciách evolúcie a akvizície jazyka (významov), pozri (Takáč 2006a-c, 2007a, b).

2 Logika a usudzovanie

V európskej kultúre sa už po tisícročia kultivuje(ú) model(y) usudzovania v termínoch logiky. Tieto modely explicitne alebo implicitne rozvíjajú predstavy o korektnej argumentácii (rozvíjajú rozlišovacie kritérium korektnej argumentácie). V podstate ide o poňatie korektnej argumentácie, s ktorým pracoval (napríklad) už Sokrates. V Platónových dialógoch vystupujúci Sokrates ukazoval svojim partnerom (oponentom) v dialógu v hlavných rysoch toto: ak by som uvažoval o doméne (probléme) X presne tak, ako vy uvažujete o doméne (probléme) Y, dospel by som k očividne nepravdivým záverom; teda – spôsob, ako uvažujete, nie je korektný, nezaručuje pravdivé závery, a preto ani vaše závery o doméne (probléme) Y nemožno akceptovať.

Takéto rozlíšenie korektnej argumentácie sa opiera o pojem kontrapríkladu⁴. Upresnenie pojmu kontrapríkladu sa dá pohodlne urobiť v jazyku, ktorý abstrahuje od konkrétností obsahu rozmanitých domén a fixuje iba *formálnu kostru* jazykov, vhodných na opis ľubovoľnej domény. Jazyk tohoto typu budeme bez ďalšieho upresňovania nazývať jazykom logiky (iba poznamenáme, že ich je veľa).

Ak teda chceme opisovať usudzovanie (operácie nad reprezentáciami) tak, aby sme mohli využívať konštrukt kontrapříkladu, je celkom prirodzené uvažovať o reprezentáciách sformulovaných, zaznamenaných v jazyku logiky. Tak sa nad bazálnou kognitívnou vrstvou reprezentácie, reprezentácie pomocou rozlišovacích kritérií (do tejto úrovne patrí aj rozlišovacie kritérium korektného argumentu) dá nastaviť poňatie reprezentácie v logickom jazyku (kde sa kritérium korektnej argumentácie nejakým spôsobom upresňuje). Kvôli pohodlnému vyjadrovaniu môžeme hovoriť o logickej reprezentácii (s poznámkou, že ide o triedu reprezentácií).

Logická reprezentácia umožňuje (okrem iného) spojiť v jednom rámci poznatky reprezentované explicitne s poznatkami, reprezentovanými implicitne (vďaka sile úsudku, opísanej presnými kritériami logického odvodzovania). Sémantickú verziu takejto reprezentácie vytvoril/dotiahol Tarski (1933) do tvaru, ktorý sa používa dodnes.

2.1 Matematická logika ako model usudzovania

Matematická logika je kánon ako robiť logiku. Každý seriózný pokus o príspevok do nejakého druhu logiky sa opiera o štandardy a metódy, vyvinuté v matematickej logike od konca 19. storočia. Usudzovanie je v matematickej logike modelované formalizáciou pojmu dôkazu. Alebo presnejšie: ak niekto chce matematickú logiku považovať za model usudzovania, jadrom toho modelu je pojem formalizovaného dôkazu.

Jednou z takýchto formalizácií sú hilbertovské systémy. Dôkaz je v nich postupnosť formúl (jazykových výrazov, vyhovujúcich jednoznačným syntaktickým kritériám), pričom každá z týchto formúl je buď axiómou alebo sa získala z predchádzajúcich formúl pomocou pravidiel odvodenia. Pravidlá odvodenia sú formulované čisto syntakticky: z výrazov určitého, presne definovaného, tvaru možno odvodiť výrazy presne definovaného tvaru. Každá formula, dokázateľná z axióm lo-

⁴ V matematike patrí hľadanie kontrapríkladov do základnej metodologickej výbavy, medzi základné spôsoby matematickej argumentácie: ak nájdeme kontrapríklad k nejakému matematickému tvrdeniu, toto tvrdenie je nepravdivé. Podľa Stenninga a van Lambalgen (2007, str. 33) to však nebolo vždy takto. V osemnástom a devätnástom storočí sa stávalo, že objavený kontrapríklad k nejakému tvrdeniu sa považoval za prípad, ktorý je mimo domény daného tvrdenia. Odvolávajú sa pritom na (Lakatos 1976).

giky, môže slúžiť ako pravidlo odvodenia matematických (alebo akýchkoľvek) tvrdení z iných matematických (akýchkoľvek) tvrdení. K čisto syntaktickému konceptu dôkazu (odvodenia) existuje sémantický (teoreticko-modelový) pojem vyplývania. Popri hilbertovských systémoch existujú aj iné typy prístupu k formalizácii pojmu dôkazu (prirodzená dedukcia, sekventy, tablá), pozri napr. (Kvasnička 2007), kde čitateľ môže nájsť aj diskusiu z hľadiska kognitívnej vedy.

Nemá veľký zmysel priesa o to, či matematická logika poskytuje model usudzovania. Tento model však nemodeluje ani usudzovanie matematikov (v dostatočnom rozsahu). Matematickú logiku možno považovať za idealizáciu (model) prezentácie/komunikácie matematických výsledkov (Barwise, Etchemendy 1998). V tomto modeli je v centre pozornosti ideálny koncept dôkazu, kontrolovateľného mechanicky. Popri mechanickej kontrolovateľnosti treba spomenúť aj to, že tento model umožňuje aj metódy automatického dokazovania.

Zastavme sa pri tom, že v modeli usudzovania, ktorý ponúka matematická *logika*, sa dajú formulovať pravidlá korektného usudzovania. Každý teorému logiky (vždy pravdivej formule logiky) zodpovedá nejaké pravidlo korektného usudzovania. Takýmito pravidlami sú napríklad modus ponens (z výrokov tvaru A, ak A tak B možno odvodiť výrok tvaru B) alebo odstránenie konjunkcie (z výroku tvaru A a B možno odvodiť výrok tvaru A).

Podobne aj rôzne neklasické logiky (mnohohodnotové, intuicionistická, modálne, epistamické atď.) formulujú a dokazujú teorémy, platné v ich systémoch a tieto teorémy môžu slúžiť ako pravidlá usudzovania. Nechcem tvrdiť, že takéto pravidlá človek vôbec nepoužíva (vedome či nevedome). Nevie si však dobre predstaviť, že by nejaký človek usudzoval napríklad s použitím pravidla odstránenia konjunkcie (samozrejme, pokiaľ nie je na cvičení z logiky).

Na druhej strane, človek veľmi často (možno spravidla) usudzuje tak, že sa to nedá charakterizovať pomocou nejakej logickej formule. Môj mladší vnuk Michal niekedy v treťom roku jeho života prekvapil protiotázkou: “Načo sa pýtaš, keď to vieš?”. Určite nevychádzal z nejakých premís a neodvodil záver podľa nejakého pravidla. Podstatné je však to, že sa jeho usudzovanie ani nedá zrekonštruovať takým spôsobom. Bez akýchkoľvek pochybností ho však k tejto otázke viedol nejaký úsudok, ktorý pozostával z implicitnej a intuitívnej formulácie nejakého princípu (povedzme, že pravidla racionálnej komunikácie), ďalej zo zistenia nesúladu medzi týmto princípom, otázkou, ktorú dostal a jeho znalosti o tom, že pýtajúci sa pozná odpoveď na otázku. Preto protiotázkou vyjadril, že na neracionálnu otázku nemieni odpovedať.

Veľmi by sme ochudobnili poznanie ľudského usudzovania, keby sme odmietli venovať pozornosť rozmyšľaniu takéhoto (a ďalšieho) druhu. Inými slovami, keby sme odmietli hľadať ďalšie modely usudzovania.

Častým typom usudzovania je približne tento: Máme dve neoverené domnienky. Tieto domnienky sú nekonzistentné s ostatnými našimi znalosťami, ak by mali platiť súčasne. Ak niektorú z týchto domnienok preferujeme, akceptujeme ju a druhú nie (aspoň predbežne). Opäť, tento úsudok nie je založený na nejakom pravidle, kopírujúcom nejaký vždy pravdivý výrok logiky. Tomuto spôsobu usudzovania sa budeme venovať nižšie, v časti o nemonotónnych logikách.

Uvedené príklady ukazujú, že výskum (modely) usudzovania možno a treba robiť rozmanitým spôsobom, často veľmi odlišným od prístupu kanonizovaného matematickou logikou. Mnohé z týchto spôsobov sú zaujímavé aj z hľadiska široko poňatej logiky, pozri napr. Makinson (2003), sú tu však aj ďalšie možnosti.

2.2 Nemonotónne logiky ako modely usudzovania

Matematická logika je monotónna v tomto zmysle slova: ak k nejakej množine axiém (predpokladov) pridáme ďalšie axiémy (predpoklady), z rozšírenej množiny odvodíme všetko, čo sme odvodili z pôvodnej (a v typickom prípade aj niečo navyše).

Ľudia však zvyknú rozmyšľať (usudzovať) aj inak – ak získajú novú informáciu⁵, odmietnu niekedy niektoré závery, ktoré odvodili z pôvodnej (pod)množiny informácií. Takéto usudzovanie je nemonotónne: závery z podmnožiny informácií netvoría podmnožinu záverov z nadmnožiny informácií.

Nemonotónnosť je charakteristická pre hypotetické usudzovanie. Ak závermi usudzovania sú nejaké hypotézy, sme pripravení hypotézy odmietnuť, ak sú v rozpore s faktami alebo lepšie overenými (viac preferovanými) hypotézami. Niektoré druhy takéhoto usudzovania (prirodzene súvisiaceho s revíziami) študujú nemonotónne logiky. Nemonotónne logiky predstavujú zaujímavý a dôležitý prístup k reprezentácii znalostí a usudzovaniu. Azda dominantným nemonotónnym formalizmom súčasnosti sú logické programy so sémantikou stabilných modelov (answer set). Tento formalizmus nie je iba predmetom teoretického záujmu, ale slúži aj ako implemetačný nástroj (v rámci paradigmy answer set programming). V nemonotónnej logike nie je v centre pozornosti dôkaz, ale sémantika. Nemonotónne usu-

⁵ Používam tu slovo informácia, ale mám na mysli stále nejaké výroky, dokonca nimi môžu byť aj formuly akéhosi logického jazyka.

dzovanie sa neopiera o pravidlá, kopírujúce dokázané tvrdenia. O čo ide pri súčasných formalizáciách nemonotónneho usudzovania si ilustrujeme na príklade stabilnej sémantiky.⁶

Logický program je množina pravidiel tvaru

ak platí ... potom platí ...,

pričom medzi predpokladmi takýchto pravidiel môžu byť aj takzvané defaultové negácie tvaru $\text{not } A$ (intuitívne, $\text{not } A$ je pravdivé, ak nie je známe, že by bolo pravdivé A). Vidno, že defaultové negácie majú fakticky status hypotéz. Kvôli jednoduchosti budeme uvažovať iba o normálnych logických programoch – používajú iba také pravidlá, kde dôsledkom je jeden atóm daného jazyka.

Pojem stabilného modelu logického programu P si môžeme priblížiť takto: Pridajme k P množinu defaultových negácií (hypotéz) H . Zo zjednotenia H a P odvodíme množinu dôsledkov (atómov) D . Nech N je množina všetkých atómov, ktoré nepatria do D . Ak do H patria presne tie $\text{not } A$, kde A patrí do N , potom D je stabilným modelom programu P . Intuitívne, vtedy možno H považovať za maximálne nasýtenú množinu hypotéz, ktoré spolu s programom P dávajú zmysel.

Formálne sa takéto sémantiky definujú tak, že sa skonštruuje nejaký operátor a sémantická charakterizácia daného nemonotónneho formalizmu (napríklad stabilný model) je pevným bodom takéhoto operátora (takýmto sémantikám sa hovorí fixpointové).

Nemonotónne logiky určite predstavujú nové a zaujímavé modely usudzovania. Keďže týchto modelov je viac, nemožno jedným dychom charakterizovať ich funkciu. Vo všeobecnosti však možno povedať, že vytvárajú idealizované modely s využitím prostriedkov, kultivovaných matematickou logikou. Tieto modely majú ambíciu dať buď nejaký globálny pohľad na množiny všetkých vzájomne zlučiteľných hypotéz a faktov alebo charakterizovať nejaký vzťah nemonotónneho (minimálneho, preferenčného apod.) vyplývania.

Konkretizujme si to na príklade stabilnej sémantiky. Stabilný model je produktom nejakej konštrukcie, ktorá je globálna v tomto zmysle slova: o každom atóme je rozhodnuté, či je pravdivý, o každej hypotéze je rozhodnuté, či môže byť pravdivá vzhľadom na daný stabilný model.

Idealizované modely nemonotónnych logík sa líšia od toho, ako človek rozmýšľa. Človek, keď rozmýšľa hypoteticky, rozmýšľa viac lokálne, neobsiahne celé spektrum možných hypotéz a faktov.

⁶ samozrejme, nemonotónne logiky nie sú jednoliaty, nediferencovaný celok, preto si na tomto priestore nemôžem dovoliť všeobecne platnú charakterizáciu. Sémantika stabilných modelov však predstavuje dôležitú, nie okrajovú, paradigmu.

3. Alternatívy k logike

Budeme sa teraz venovať pokusom o *znovupochopenie* usudzovania, inými slovami niektorým neštandardným modelom usudzovania.

3.1 Heterogénne usudzovanie

Zaujímavá skúsenosť získali Barwise a Etchemendy, logici zo Stanfordu, pri používaní softvéru⁷ ako didaktického nástroja pri kurzoch logiky. Práca s ním výrazne zlepšila efektívnosť vyučovania logiky a výkon študentov. Najväčší a neočakávaný prínos však spočíval v tom, že študenti používali rôzne netradičné (z hľadiska logiky) postupy, ktoré viedli designérov toho softvéru Barwisa a Etchemendyho k zmene pohľadu na povahu logiky. Počítač zrevolucionizoval ich chápanie najzákladnejších pojmov logiky (Barwise, Etchemendy 1998).

Softvér (Turing's world, Tarski's world) používa popri propozícnej aj grafickú reprezentáciu, vizualizáciu. Možno hovoriť o heterogénnej reprezentácii. Barwise a Etchemendy (1998) konštatovali, že vďaka nej sa študenti pri usudzovaní, riešení problémov viac sústredili na obsah než na syntaktickú stránku. Navyše a podstatne, ich usudzovanie (vtedy, keď bolo korektné i vtedy, keď bolo nekorektné) prebiehalo inak, než by ich učitelia očakávali. Teória usudzovania, ktorú chceli študentov učiť, sa ukázala byť neadekvátnou na vysvetlenie a postihnutie toho, ako študenti v skutočnosti usudzovali. Pokus o transformáciu grafickej reprezentácie do reprezentácie v tvare viet vedie k zmene jednoduchého dôkazu v dôkaz pozostávajúci zo stoviek krokov a v ktorom sa hlavná myšlienka pôvodného dôkazu zatmnie. Pod vplyvom týchto skúseností si Barwise a Etchemendy začali hlbšie uvedomovať význam všeobecne známeho (ale v modeloch usudzovania nedostatočne zohľadňovaného) faktu, že ľudia v každodennom i vedeckom usudzovaní používajú a kombinujú informácie vyjadrené v mnohých formách (množiny viet nejakých jazykov, tabuľky, rôzne grafy, náčrty, vizualizácie, mapy atd.).

Výsledkom tohto prehodnocovania a premýšľania bol tretí softvérový produkt (courseware), nazvaný Hyperproof (pozri citovanú linku na web), pomocou ktorého začali učiť študentov metódy heterogénneho usudzovania, kombinujúceho operácie na vetnej (propozicnej) a grafickej reprezentácii.

Ešte je predčasné hovoriť o uzavretej teórii (modeli) heterogénneho usudzovania, predsa však Barwise a Etchemendy načrtli isté základné črty. Kľúčovými pojmami tradičného modelu usudzovania, založeného na logike sú pojmy dôkazu a protipríkladu. Tieto pojmy sa

⁷ Pozri <http://www-csli.stanford.edu/hp/>

však vzťahujú na výsledky (závery) usudzovania, nie na samotný proces usudzovania. V modeloch založených na logike nie je postihnutá kľúčová zložka procesu usudzovania – získavanie evidencie/argumentov v prospech/proti nejakému záveru.

Ambíciou modelu heterogénneho usudzovania je poskytnúť metódy usudzovania aj pre tie prípady, keď usudzovanie neskončí dôkazom alebo vyvrátením nejakého tvrdenia, dokonca aj pre tie prípady, keď nie je známe, k akému záveru usudzovanie dospeje (tak to býva často pri každodennom i vedeckom usudzovaní). Dôležitým komponentom tu sú metódy zisťovania relevantných informácií (exploration of a space of possibilities).

3.2 Model zvieracieho usudzovania

Jedným z priekopníkov logického modelovania nemonotónneho usudzovania bol John Mc Carthy (1980), Výzvu modelovať toto usudzovanie prostriedkami, charakteristickými pre modernú logiku spojil so sloganom „Jumping to conclusions“. Vyjadril tým nádej, že posun od modelovania dôkladnej dedukcie k modelovaniu takého usudzovania, pri ktorom sa neberú do úvahy všetky možné interpretácie jazyka, umožní vytvoriť systémy (modely usudzovania), ktoré sú z výpočtového hľadiska efektívnejšie a obrazne umožnia skákať k záverom. Žiaľ, systematická formalizácia nemonotónneho usudzovania viedla k problémom, ktoré sú z výpočtového hľadiska ešte ťažšie ako výpočtové prolémy, spojené s dokazovaním a splniteľnosťou v rámci matematickej logiky.

Stále však ostáva výzvou opísať/modelovať ľudské usudzovanie, ktoré skutočne „skáče k záverom“ (Šefránek 2007). Človek, keď usudzuje, málokedy robí dlhé reťaze odvodení. Veľmi často po pár krokoch „vidí“ správny záver, niekedy sa však mylí. Model kombinujúci takéto rýchle, ale omylné usudzovanie s logicky korektným, ale výpočtovo náročným usudzovaním možno považovať za (zatiaľ nedosiahnutý) cieľ, ktorý môže prispieť k hlbšiemu pochopeniu usudzovania i k jeho efektívnym výpočtovým implementáciám.

Dá sa predpokladať, že analýza zvieracieho usudzovania pomôže priblížiť sa k tomu cieľu. Domnienka, že schopnosť rýchleho usudzovania má evolučné korene a že jej opísanie a pochopenie by sa mohlo opierať o opísanie a pochopenie zvieracieho usudzovania, sa zdá byť sľubná, hodna preverenia. Pochopenie a opísanie nejakého javu v tej najjednoduchšej podobe je z metodologického hľadiska dôležité pre pochopenie a opis jeho komplikovanejších variant.

Samozrejme, ako základnú otázku možno postaviť, či má zmysel hovoriť o zvieracom usudzovaní. Mnohé experimenty a pozorovania

(minimálne) naznačujú, že usudzovanie možno pripísať aj vyšším živočíchom. Úsilie nájsť také sémantické pojmy, ktoré umožnia zmysluplne hovoriť o usudzovaní zvierat možno považovať za produktívne. Ak nájdeme takýto sémantický opis, mohli by sme ho aplikovať na jednoduchých (abstraktných) agentov bez toho, že by sme u nich predpokladali mentálne procesy.

Autori experimentu opísaného v (Bräuer 2006) testovali schopnosť zvieratá (šimpanza a domestikovaného psa) rozpoznať, kde je ukrytá potrava. Toto rozpoznanie mohlo využívať tri typy podnetov: komunikatívne, kauzálne a behaviorálne. Zameriame sa na prvé dva typy, ktoré sú pre tento experiment podstatné.

Úlohou pokusného zvieratá bolo vybrať si z dvoch nádob obrátených hore dnom tú, pod ktorou sa nachádza potrava. Ako komunikatívne podnety sa použili ukazovanie na nádobu s potravou a pohľad smerom na nádobu. Kauzálne podnety boli realizované fyzickou manipuláciou (napríklad hrkaním nádobou). Ak zviera rozumelo kauzálnym vzťahom medzi lokalizáciou zvuku a umiestnením potravy, mohlo na tom základe usudzovať.

Šimpanzy veľmi dobre reagovali na kauzálne podnety, no nevenovali veľkú pozornosť ukazovaniu. Psy naopak spoľahlivo nasledovali komunikačné podnety, ale umiestnenie potravy neboli schopné odvodiť na základe kauzality.

Podľa citovanej práce (a mnohých ďalších) má zmysel hovoriť o tom, že zvieratá **rozumejú a usudzujú, odvodzujú dôsledky z toho, čo pochopili**. Opice rozumejú kauzálnym podnetom, psy komunikačným.

Vedy, ktoré sa tradične venujú skúmaniu usudzovania, chápania (a súvisiacich fenoménov ako je *poznatie* a *význam*) sa sústreďujú na poznatky a úsudky, sformulované v *jazyku*. Veľkou výzvou (aj pre tieto vedy) je skúmať, čo je usudzovanie a pochopenie ako *biologický fenomén*, ktorý sa vyskytuje v prírode v *predjazykovom štádiu*.

Verím, že takto postavený cieľ pomôže hlbšie, mnohotvárnejšie a v inom svetle pochopiť ľudské usudzovanie, poznanie, reprezentáciu poznania a významov.

V (Retová et al. 2007) sme načrtli pojmovú aparatúru, v ktorej sa dá upresniť (modelovať) predstava o usudzovaní a chápaní agentov s predjazykovým správaním. Použili sme sémantiku rozlišovacích kritérií.

Pochopenie stotožňujeme s *rozpoznávaním* a *tvorením významov (rozlišovacích kritérií)*. Ak organizmus (agent) dokáže používať nejaké rozlišovacie kritérium pri rozpoznávaní objektov, vlastností (atd.), alebo keď vytvára nové rozlišovacie kritériá, aby sa zoriento-

val v prostredí, môžeme hovoriť, že chápe (vlastnosť, vzťah, situáciu, udalosť, závislosť, pravidlo atd.).

Experimenty z (Bräuer et al. 2006) sme teda analyzovali v sémantickom rámci rozlišovacích kritérií. Definovali sme rozlišovacie kritériá situácií, udalostí, typov situácií, typov udalostí. problémov a pravidiel.

Pravidlá sú špeciálne rozlišovacie kritériá, ktoré sa používajú pri usudzovaní. Nejakým rozlišovacím kritériám typov situácií alebo udalostí (významom predpokladov pravidla) priradia nejaké rozlišovacie kritériá (významy záveru).

Zaviedli sme dva druhy pravidiel – akčné a situačné. Dôsledkami prvého z nich sú rozlišovacie kritériá akcie, u druhých rozlišovacie kritériá typu situácie. Intuitívne, prvé vedú agenta ku konaniu, druhé k rozpoznaniu stavu prostredia.

Pri tejto sémantickej konštrukcii sa ukazuje, že usudzovanie je špeciálnym prípadom chápania: ak niekto alebo niečo dokáže produktívne usudzovať, v istom zmysle slova chápe vzťahy (napríklad medzi udalosťami).

Ako možno predpoklad o tom, že opice (alebo iné vyššie živočíchy) usudzujú, potvrdiť pozorovaním, t.j. ako možno operacionalizovať hypotézu o usudzovaní opíc (alebo iných živočíchov)? Inú možnosť ako cez pozorovanie správania nemáme. Pozorujeme správanie opíc (alebo ľubovoľných agentov v predjazykovom štúdiu) v určitých (takmer) konštantných podmienkach počas udalosti toho istého typu. Predpokladajme, že za týchto podmienok má opica na výber z rôznych možných správání. Ak v štatisticky významnej miere si vyberá jedno z nich a jej rozhodovanie sa zmení pri určitej zmene daných podmienok, môžeme prijať, že daný výber správania sa deje na základe usudzovania. Dôležité je dodať, že presnejšia špecifikácia bude s najväčšou pravdepodobnosťou závislá na konkrétnej doméne a konkrétnom rozhodovaní (podobne sú aj defaultové pravidlá doménoovo špecifické; mimochodom, v (Retová et al. 2007) sme použili aj trochu komplikovanejšie typy pravidiel, ktoré úplne zodpovedali defaultovým pravidlám).

Konceptuálny aparát rozlišovacích kritérií je flexibilný a možno ho použiť od opisu veľmi jednoduchých prípadov rozlišovania až po rozlišovanie, opreté o používanie syntakticky bohato štruktúrovaného jazyka, charakterizovateľného teoreticko-modelovou (alebo komplikovanejšou) sémantikou. Domnievame sa, že analýza, formalizácia a implementácia rýchleho (odlišného od logicky korektného, ale zdĺhavého) usudzovania v jazyku môže významne ťažiť z pokusov pochopiť zvieracie usudzovanie. Na rozlišovacie kritériá typov situá-

cií, udalostí pravdepodobne možno naviazať rozmanité, heterogénne typy reprezentácií (od propozičných po grafické vizualizácie). Od dôkladného štúdia zvieracieho usudzovania možno urobiť dôležité kroky pre pochopenie špecificky ľudského usudzovania (a predpokladáme, že tak možno dosiahnuť podstatne mnohotvárnejšie charakterizácie, než sú tie, ktoré dnes študuje logika).

Jednou z dôležitých okolností, ktoré pravdepodobne umožňujú rýchle usudzovanie, sú evolúciou zakódované preferencie pre rozhodovanie a usudzovanie. Takéto preferenčné usudzovanie (ktoré sa, mimochodom, študuje aj v kontexte nemonotónnych logík) možno založiť aj na reprezentácii pomocou rozlišovacích kritérií, menovite pomocou pravidiel, ktoré sú “pridrátované”, nejakým spôsobom pripojené, k rozlišovacím kritériám situácií a udalostí. Takto poňaté preferenčné usudzovanie môže realizovať aj dynamické preferencie – k rôznym fázam priebehu nejakej udalosti môžu byť pripojené iné pravidlá.

4. Závery

Pokúsil som sa v istom náčrte opísať niektoré modely usudzovania. Všetky boli opreté o akýsi typ reprezentácie. Domnievam sa, že tieto (a ďalšie reálne alebo možné) modely a ich reprezentácie, ukazujú, že hlbší zmysel má uvažovať a skúmať celé spektrum typov reprezentácií, než hovoriť o protipóloch symbolovej⁸ a subsymbolovej reprezentácie.

Napriek tomu, že hlavný dôraz som položil na pojednanie o modeloch usudzovania, ktoré sa líšia od klasického modelu logiky, som presvedčený, že logika (a sémantika) v širokom zmysle je fundamentálna disciplína kognitívnej vedy. Charakteristickou črtou kognícií je totiž (viac či menej) úspešná adaptácia na prostredie (poznatie prostredia). Preto pre opis kognícií treba zamerať pozornosť aj mimo nich (externalisticky) a používať konceptuálny aparát, vhodný na pochopenie kognitívnych schopností, potrebných pre poznanie prostredia. Tento pojmový aparát v tej či onej podobe potrebuje pojem pravdy a potrebuje opisovať operácie pravdu objavujúce alebo zachovávaajúce. V inej terminológii sa dá hovoriť, že sú potrebné pojmy na opis spracovania informácií kognitívnymi systémami (Sterling, van Lambalgen 2007). Mimochodom, Sterling a van Lambalgen vo svojej knihe ukazujú, ako pomocou pojmového aparátu logiky možno adekvátnejšie opisovať dáta, získané psychológmi (v porovnaní s tým ako to niekedy robia psychológovia).

⁸ Osobne viac rozumiem označeniu **propozičná** reprezentácia, než mnohovýznamovému “symbolová reprezentácia”.

Pod logickými metódami a pojmovým aparátom nemám na mysli nejakú konkrétnu logiku, ale tendenciu vytvárať formálne, schematické modely (teda, matematický opis) usudzovania a reprezentácie. Tieto modely podstatným spôsobom používajú pojmy pravdy, významu, korektnosti (alebo im blízke). Tieto metódy a tento pojmový aparát treba používať pri úsilí pochopiť kognície aspoň v tom rozsahu ako modely (pojmy a metódy), ktoré sú založené na istých typoch automatov, sietí, mechanizmov evolúcie a ktorých výstupom sú simulácie.

PodĎakovanie: Tento príspevok vznikol za podpory grantovej agentúry VEGA v rámci grantovej úlohy 1/3112/06.

Literatúra

Barwise, J., Etchemendy, J. (1998) : Computers, visualization, and the nature of reasoning. In: *The Digital Phoenix: How Computers are Changing Philosophy* (T.W. Bynum, J.H.Moor, eds.), pp. 93–116

Bräuer, J. et al. (2006): Making Inferences about the Location of Hidden Food: Social Dog, Causal Ape. In: *Journal of Comparative Psychology*, vol. 120, No. 1, 38-47

Gärdenfors, P. (1995) : Cued and detached representations in animal cognition. In *Behavioural Processes*, Volume 35, Issues 1 – 3, December 1995, Pages 263-273

Fodor, J. (2000): *The mind doesn't work that way*, The MIT Press.

Kováč, L. (2004): Komentovaný úvod do kognitívnej biológie. Zborník seminára *Kognice a umelý život IV*.

Kováč, L. (2006): Princípy molekulárnej kognície. Zborník seminára *Kognice a umelý život VI*.

Kvasnička, V. (2007): Mentálne modely logiky v kognitívnej vede. *Kognice a umelý život VII*.

Lakatos, I. (1978): *Proofs and refutations*. Cambridge University Press.

Lúčny, A. (2007): Modelovanie na báze reaktívnych agentov. In Kvasnička, V., Trebatický, P., Pospíchal, J., Kelemen, J.: *Mysel, inteligencia a život*, Vydavateľstvo STU, Bratislava.

Makinson, D.(2003) : Ways of doing logic. What was different about AGM 1985. *Journal of Logic and Computation*.

McCarthy, J. (1980) : Circumscription-A Form of Non-Monotonic Reasoning. In: *Artificial Intelligence*, vol, 13, pp. 27-39

Retová, D., Šilliková, J., Šefránek, J. (2007): Opice, psy. símantika a logika. In: *Kognice a umelý život VII*.

Stenning, K., van Lambalgen, M. (2007): *Human reasoning and cognitive science*. To be published by MIT Press.

Šefránek, J. (2002) : Kognícia bez mentálnych processov. In J.Rybár, L. Beňušková, V. Kvasnička (eds.): *Kognitívne vedy*. Kaligram, Bratislava.

Šefránek, J. (2007) : Kognitívna teória usudzovania. In Kvasnička, V., Trebatický, P., Pospíchal, J., Kelemen, J.: *Mysel, inteligencia a život*, Vydavateľstvo STU, Bratislava.

Takáč, M., (2006a) : Cognitive Semantics for Dynamic Environments. In: Hitzler, P., Schärfe, H., P. Øhrstrøm (eds.): *Contribution to ICCS 2006 - 14th International Conference on Conceptual Structures*. Aalborg University Press, Aalborg, Denmark, pp. 202-215.

Takáč, M. (2006b) : Kognitívna sémantika rozlišovacích kritérií. In: Kelemen, J., Kvasnička, V. (eds.): *Kognice a umělý život VI*. Slezská univerzita, Opava, pp. 363-372.

Takáč, M. (2006c) : Categorization by Sensory-Motor Interaction in Artificial Agents. In: Fum, D., Del Missier, F., Stocco, A. (eds.): *Proceedings of the 7th International Conference on Cognitive Modeling*. Edizioni Goliardiche, Trieste, Italy, pp. 310-315.

Takáč, M. (2007a) : Autonomous Construction of Ecologically and Socially Relevant Semantics. *Cognitive Systems Research* (in press).

Takáč, M. (2007b) : Kognitívna sémantika komplexných kategórií založená na rozlišovacích kritériách. In: Kvasnička, V., Trebatický, P., Pospíchal, J., Kelemen, J.: *Mysel, inteligencia a život*, Vydavateľstvo STU, Bratislava.

Tarski, A. (1933): *Pojęcie prawdy w językach nauk dedukcyjnych*. Warszawa: Nakładem Towarzystwa Naukowego Warszawskiego.

Vnímání z pozic konstruktivismu a enaktivního přístupu (Jak to vidí dva potkani v Morrisově vodním bludišti)

Radovan Šikl¹

Experimentátorovy ruce vložily do bazénku s vodou dva potkany. Potkani šlapou vodu a rozhlíží se kolem.

A: Plavete rád?

B: Ani ne. A vy?

A: Já taky ne. Vidíte ty značky?

B: Jasně, jsou tu jako vždycky.

A: Máme najít ostrůvek. Pak to skončí.

B: Já vím. Nemáte tušení, proč nás sem tentokrát dali dva?

A: Nevím. Asi nějaký jiný experiment.

B: Asi ano. Možná sledují, jak budeme spolupracovat nebo tak něco.

A: Co budeme dělat?

B: To co vždycky. Najdeme ostrůvek.

Ostrůvek s potravou je pod vodní hladinou a potkani jej nevidí. Mohou se ovšem orientovat podle značek na okraji bazénku. Značky jsou tři a tvoří rovnostranný trojúhelník, v jehož těžišti se ostrůvek nachází. To oba potkani dobře vědí. Zatím ale šlapou vodu a přemýšlí.

A: Máte nápad, jak to provést? Chtít po potkanovi, aby měl po celou dobu hledání potravy v hlavě mapu prostoru, pod kontrolou informace o poloze cílového bodu a o svém směřování, aby bezchybně odhadoval a srovnával vzdálenosti, mi při našem mizerné zraku případně dost těžký, skoro až nadpotkaní úkol. Neznáte náhodou nějakou účinnější strategii?

B: Znímám, ale trochu hůře se vysvětluje. Narýsovat si v mysli schématickou mapu prostoru – sít' vzdáleností, směrů, pozic, kót – lze celkem snadno. Chovat se podle ní už jde ale o dost hůř. Vlastně si nemyslím, že se tento postup pro naši úlohu hodí a sám raději řeším zadání úplně jinak. Víím, že hledané místo je někde mezi vrcholem trojúhelníka a středem dvou protilehlých vrcholů, tedy na kolmici, a na té kolmici o něco blíž k základně než k vrcholu. To mi stačí, žádnou jinou informaci k vyřešení úlohy už nepotřebuji. Z výchozího stanoviště – pokud to právě není jeden z vrcholů – se kolmice

¹ Psychologický ústav AV ČR, Veveří 97, 602 00 Brno, e-mail: sikl@psu.cas.cz

vytyčuje špatně, a tak raději nejprve poplavu k nejbližší značce a odtud pak směrem k protilehlé straně, přičemž obě značky před sebou se budu snažit udržovat stále ve stejné vzájemné vzdálenosti. Někde za polovinou cesty zvýším ostražitost. Víím, že cíl musí být blízko. A většinou se nespletu. Rozumíte?

A (*podrbe se za uchem*): No snažím se, ale zatím mi uniká, co vlastně z celého bludiště vidíte nebo jak ho vidíte. Vždyť se přeci primárně jedná o percepční úlohu.

B: No to ano.

A: A říkáte, že dbáte jen, aby ty dva vrcholy byly neustále ve stejné vzdálenosti a že v jednu chvíli začnete dávat víc pozor. Promiňte, ale to mi přijde vzhledem ke složitosti terénu jako dosti slaboduchá strategie.

B: Hm?

A: Tedy ve srovnání se mnou.

B: Poslyšte... Vnímání není nějaký obrázek odstřihnutý od našeho fungování v realitě, naopak vnímání se odráží v našem jednání. Pohyb v prostředí je namnoze přímo řízený zrakem (případně jinými smysly) a přesnost, účelnost či komplexnost tohoto pohybu vypoovídají o úrovni vnímání. Doplnuji, že k jednání žádné mapy v mysli nejsou potřeba.

A (*ohromeně*): Opravdu?

B: Spousta zvířat je výtečná v navigaci, aniž by si vytvářela představu o prostoru, ve kterém se pohybují. Rozumí dobře vztahu mezi vlastním pohybem a měnící se senzoricou stimulací a to jim ke složitým manévřům stačí. Motýlí samec v bezprostředním závěsu kopíruje trajektorii samičího letu při námluvách. Pelikán se střemhlav vrhá do vody za potravou a ve velké rychlosti sklápí křídla tak, že ve chvíli ponoru už je má složená u těla. To jsou úchvatné doklady skvělého prostorového chování, které se ovšem bez přeměrování a vyhodnocování prostorových vztahů obejde.

A (*zaraženě*): Ale jak potom tuto schopnost vysvětlíte?

B: Živočich v podobných situacích svůj pohyb prostorem řídí podle jednoduché optické veličiny, jejíž hodnotu se snaží udržet na konstantní hodnotě. Nic jiného. Neděste se, uvedu příklad. Přistávající včela mění rychlost svého letu tak, aby obraz plochy, kam dosedá, v jejím oku, resp. tempo zvětšování tohoto obrazu zůstávalo stejné. Čím je tedy blíže místu přistání, tím je její let pomalejší. Výlučným soustředěním se na jeden znak, nadto takto simplexní, si pohybovou úlohu výrazně usnadňuje. Nemá neadekvátní ambice všechno vidět, nepodstatností si nevšímá.

- A (*ožije*): Počkejte, to já si ale také nemyslím, že pozorovatel s sebou neustále nosí úplný obraz prostředí, v němž se právě pohybuje. Nicméně to ještě neznamená, že zvířata si při pohybu žádné představy nevytvářejí, ani že představy nemohou jednání významně ovlivnit. Mohu vám uvést příklady prostorových odhadů, které jsou dobrou výchozí informací pro pozdější pohyb. Jak jistě víte, my potkani, chceme-li v kanálu přeskočit stoku, nejprve začneme vydávat píšťivé zvuky o vysoké frekvenci směrem k protějším břehu a teprve podle odrazu poznáme šířku vodní překážky, kterou se následně buď odhodláme, nebo neodhodláme přeskočit. Netopýři a delfini pomocí echolokace určují polohu a vzdálenost potenciální kořisti. Kobylky a sarančata si vytváří představu o vzdálenostech z parallaxy pohybu – kývou hlavou ze strany na stranu a srovnávají rychlost přemísťování různě vzdálených objektů v oku. Chřestýši a jiní hadi se při útoku řídí podle termomapy okolního prostředí.
- B (*zaujatě*): To je zajímavé, co říkáte. Zdá se, že v některých prostorově náročných situacích může být dobrá představa prostředí k užitku, v jiných je zbytečná. Souhlasil byste s myšlenkou, že při orientaci v prostoru spoléháme na integraci tras svého pohybu a jednoduché optické veličiny a mezi pohyby se můžeme pokusit aktualizovat své umístění v prostoru?
- A: To by šlo. Musím uznat, že pravdě o prostorovém chování živočichů budete blíže asi vy. U jednodušších forem kognice je pohyb prostředím obvykle plynulý a reakce na podněty okamžité. Ale úplně jiné to bude u člověka. Lidské vztahování se k prostoru je mnohem komplexnější; prostředí méně předvídatelné a pohyb v něm strukturovanější; člověk se často ocitá v situacích, kdy musí řešit vzniklý problém. Víte co, pojďme se teď už bavit raději jen o lidech a o lidském vnímání. (*potutelně*) Lidé jsou tak nějak zajímavější...
- B: Dobrá, proč ne. Ale to se pletete, jestli si myslíte, že jste tím získal pro své argumenty významnou výhodu. Vyjasněme si nejprve pozice. Popište mi svou představu o lidském vnímání.
- A (*chvilu přemýšlí, srovnává si v hlavě myšlenky*): Podstatným omezením a problémem pro vnímání je informační chudost a mnohoznačnost obrazu světa dopadajícího na sítnici oka. Přitom sítnicové obrazy jsou pro pozorovatele jedinou vstupní informací. Ta ovšem k úspěšnému vnímání rozhodně nepostačuje. Podnětovou informací je potřeba dotvořit tak, aby lépe vyhovovala struktuře lidské mysli a bylo možné ji smysluplně interpretovat. Vstupní data jsou v průběhu zpracování konfrontovaná s odpovídajícími mentálními obsahy – konkrétními vzpomínkami, příslušnými kategoriemi,

pojmy, schémata, scénáři, případně také hodnotami či emocemi, které podnět u pozorovatele vyvolává. A teprve až výsledkem této konfrontace, tohoto uzpůsobování podnětu osobě pozorovatele je obraz, kterému pozorovatel rozumí, ale který je něčím víc, než co „vidí“ naše oči.

- B (*zamyšleně se podívá na nejbližší značku*): To se mi nezdá. Uvědomte si, jaké množství příchozích informací bychom bez ustání museli zpracovávat. A to ještě každý jednotlivý podnět se může v oku zobrazit v množství rozmanitých podob. Nemáme kapacitu na to srovnávat všechny informace s předobrazem ve všech jeho možných podobách. Rozumím vaši představě o osmyslení mozaiky světelných intenzit na sítnici, ale takto je neproveditelná.
- A: Zdání klame. Do zpracování zrakových podnětů se zapojuje více než polovina mozkové kůry člověka; to je víc než u jakékoliv jiné mentální operace. Již jen z toho je patrné, jaká je „obyčejnému“ vidění věnovaná péče. To, že proces zpracování podnětu nestojí žádnou vědomou námahu, vás nesmí zmást.
- B (*nenechá se zvíkat*): Jak ale vysvětlíte, že se takové zatížení mozku neprojeví v opožděných reakcích? V životě mnohdy potřebujeme na vzniklou situaci okamžitě zareagovat a při vámi popsané složitosti zpracování bychom se taky nemuseli dočkat. Než bych si vyjasnil, že to, co se na mě žene, je plně vzrostlý, dospělý krysařík, než bych si jeho počínání dal do souvislosti se svými dosavadními zkušenostmi s krysaříky, než bych nechal svůj dojem utvářet všemi emocemi, které ve mně jeho pohyby vzbouzí, bylo by už dávno po mně.
- A (*usmívá se*): No jo, to máte recht. Tedy měl byste, kdyby naše vnímání pokaždé začínalo u „nepopsané desky“ a směřovalo k jedinečnému, jen pro danou situaci platnému vjemu. Naštěstí pro nás je ale přistupování k podnětům výrazně „předpojaté“. Předně do vnímání – interpretace sítnicového obrazu – zapojujeme předpoklady o pravděpodobných podobách objektů; z mnoha možných podob podnětu odpovídajících jeho zobrazení na sítnici preferujeme právě jen některé, zpravidla podoby obvyklé a tvarově jednoduché. Podobně jsou preferované, tedy s větší pravděpodobností vnímané podoby odpovídající schématům a momentálnímu nastavení pozorovatele. Pozorovatel si zpracování usnadňuje také tím, že si nevnímá jednotlivých prvků zorného pole, ale tyto prvky automaticky seskupuje do nejpravděpodobnějších konfigurací a nadále při zpracovávání pracuje už jen s těmito celky. Výraznou měrou rovněž do zpracování zapojuje anticipaci. A abych už domluvil, uvědomte si, že naprostá většina všednodenních situací neklade na

přesnost vnímání nijak vysoké nároky, což by ostatně při lidském sklonu ke schematizaci a kategorizaci stejně nebylo možné. No snad vás taková odpověď dokáže uspokojit. Pro mě samotného je to taky trochu záhada.

B (*kategoricky*): Promiňte, ale nedokáže. Většinu z toho, co jste právě řekl, považuji za nepodstatné. Oprošťujete totiž lidské vnímání od prostředí, v němž se odehrává. Vnímající jedinec neexistuje v izolaci od okolního světa, naopak jeho vnímání je výchozím bodem zasahování do dění. Vnímání charakterizuje právě to, že je součástí explorační, jednání, aktivní existence jedince v prostředí. Pouze v tomto kontextu má smysl snažit se vnímání porozumět.

A (*nevěřicně*): Prosím?

B: Abyste mi rozuměl... Svět poznáváme prostřednictvím vlastní aktivity, tak že se pohybujeme v prostředí. Nedíváme se na svět z odstupu jako nezúčastněný pozorovatel. Naopak tím co děláme, jak jednáme, jak vstupujeme s prostředím do interakce, silně ovlivňujeme to, co uvidíme. A zpětně to, co uvidíme, usměrňuje naši pohybovou aktivitu. Vnímání je spojené s jednáním a jednání s vnímáním.

A: Počkejte. Zrak nám přeci zdaleka neslouží jen k tomu, abychom mohli jednat, ale také abychom se o světě něco dozvěděli, jednoduše proto že to chceme vědět. Lidé jsou velcí konzumenti informací. Poznávají i tehdy, když s objektem zájmu nezamýšlejí nebo ani nemohou přímo vstoupit do interakce, například při sledování televize nebo shonu za oknem kavárny. Troufnu si tvrdit, že dokonce většina podnětů je takových. Takže co se týče vztahu k jednání: vnímáním získáváme o podnětu důležité poznatky, které mohou restrukturovat stávající schémata, a tím následně ovlivnit naše jednání. Ano, utvářením představ o světě může vnímání napomáhat k řízení jednání, ale nedovedu si představit, jak by to mohlo jít bez prostřednictví reprezentací.

B (*tvrdšíjně*): A já tvrdím, že pro naši existenci ve světě žádné percepční reprezentace nejsou nutné. Tvořit v mysli obrazy světa je neekonomické a neefektivní. Proč to dělat? To je zdvojování skutečností. Nemusíme mít v hlavě obraz, abychom byli v obraze. Nemusíme mít ucelenou představu o světě, mít svět pod kontrolou, abychom mohli jednat. Stačí otevřít oči. Ke světu máme v každém okamžiku přímý přístup. Podle vás je přímé vnímání vzhledem k odlehlosti skutečné a promítnuté podoby podnětů nesmysl, ale k jednání přeci stačí registrace invariantních informací o uspořádání objektů, základních vztahových informací, které se se změnou pozice v prostoru nemění. A taková informace je v sítnicovém obrazu obsažená jednoznačně a čitelně. Jedinec tedy nemusí složitě

konstruovat, ani usuzovat, aby viděl, protože vidí přímo. Ovšem vjem – nebo jak to vlastně pojmenovat – existuje pouze v jednání.

A (*otráveně*): Zatím od vás slyším jen samé silácké a přitom prázdné proklamace. O povaze vztahu vnímání – jednání zatím nepadlo ani slovo.

B: Tak poslouchejte... Lidská schopnost vnímat je vystavená na porozumění toho, jak se senzoričká stimulace mění s pozorovatelovým pohybem. Jak se například obraz objektu průběžně zvětšuje při přibližování a deformuje, když kolem něj procházíme. Samozřejmě ne všechny příchozí informace jsou takto hrubé. V usouvzažňování senzoričkého s motorickým, ve vytváření senzomotorických znalostí dosahujeme vysoké úrovně. Dokážeme – byť nevědomě – registrovat a správně vyhodnotit minimální rozdíly v příchozích informacích. Výsledek je přitom vázán na tělesné schéma. Tvrdím, že naše vnímání je na těchto senzomotorických kontingencích vystavené.

A (*zadumaně*): Podle toho, co říkáte, se skutečně zdá, že vidění a jednání mohou být propojené více, než jsem si dosud dokázal představit, nicméně vaše tvrzení, že vnímání je tělesnou aktivitou konstituované, považuji za extrémní názor. Z užitečné informace ještě nutně nevyplývá inherentní závislost.

B: To zní jako rozdíl mezi silnou a slabou verzí. Každopádně mám radost, že začínáte chápat, že zajímat se o vnímání má smysl teprve až v kontextu osoby pozorovatele.

A (*dlouze kývá hlavou*): Zdá se, že rozdíly v našich názorech jsou v samotných východiscích, v konceptualizaci vnímání. Najdeme vůbec platformu pro názorový střet, nebo se nutně musíme míjet?

B: No jo, rozdílů je víc než dost. Už v tom se neshodneme, zda hlavním výsledkem vnímání je jedincův pohyb anebo obraz v mysli; zda vnímání vyústí v prožitek světa anebo povědomí o světě; zda prostředí obklopující jedince je popsáno tím, jak jím pozorovatel cestuje anebo nezúčastněným popisem při pohledu shora.

A: Současně já tvrdím, že mezi vstupem a výstupem, tedy mezi informacemi dopadajícími na sítnici oka a zrakem řízeným jednáním je důležitý mezistupeň, kterým je percepční reprezentace. A vy obhajujete pozici, že vnímání a jednání jsou propojené bezprostředně; vnímání plynule předchází a navazuje na jiné aktivity jedince.

B: Pro vás je vnímající subjekt systémem smysly – mozek, kdežto já k němu přistupuji jako ke komplexnímu živočichovi (ne-jen pozorovateli) zasazenému a aktivně interagujícímu v prostředí.

A (*vyprskne vodou*): Ale no tak, neurázejte. Stejně tak bych já mohl tvrdit, že pro mě je pozorovatel ušlechtilým myslitelem testujícím hypotézy a pro vás jen omezeným systémem smysly – jednání.

- B: Hm... buď jak buď jsou nároky na kognitivní zpracování podnětu viděno z vaší perspektivy značné, podle mě jsou naopak zanedbatelné, stačí se naučit správně rozumět vztahu mezi pohybem a proměnami na sítnici.
- A: Rozcházejme se rovněž v názoru na časový rozměr vnímání. Jeden tvrdí, že vjem jakožto jednotlivý a statický obraz je jen dodatečný myšlenkový konstrukt, kterým není relevantní se zabývat, druhý kontruje tvrzením, že pohyb je pro pozorovatele nic než pospojovaný sled statických obrazů.
- B: Tak a tady bychom se mohli na chvíli zastavit. Podle mě je totiž pro vnímání klíčovým pojmem změna. Aktuální podoba světa se nám zobrazuje v oku, ovšem vlivem našeho pohybu nebo pohybu odehrávajícího se v okolním prostředí se neustále proměňuje. Různé podoby sítnicových obrazů srovnáváme mezi sebou a taky ve vztahu k prováděné činnosti. Jen takto je možné dosáhnout (implicitního!) porozumění toho, jak vypadá naše prostředí, co je v něm možné a co bude následovat když... Bez změny vůbec žádnou zkušenost s prostředím mít nemůžeme. Pro své tvrzení mám pár příkladů. Opravdu jenom pár, protože stavu neměnnosti obrazu je dosti obtížné dosáhnout. Naše oči jsou třikrát až čtyřikrát za vteřinu v pohybu, čímž je přísun stále nových informací zajištěn. Pokud ale umělým zásahem přeci jen dosáhneme stabilizovaného sítnicového obrazu (k oku připevníme svítící diodu, jejíž poloha bude přesně kopírovat dráhu očních pohybů), dojde k výraznému narušení vnímání. Podnět (diodu) zobrazující se průběžně na stejném místě sítnice po krátké době přestaneme registrovat. Situaci úplné nehybnosti celého obrazu opravdu zažijeme stěží, ovšem podobný efekt – vyhasínání vjemu – bude na pozorovatele mít neměnnost byť i jen jediného podnětového parametru. Oblečení na těle a brýle na nose přestáváme brzy po nasazení pociťovat, žluté zbarvení osvětlení při pobývání v pokoji za chvíli už nevidíme jako žluté. Neměnnost je pro náš poznávající organismus natolik nepřírozená, že ten v krátké době ztrácí k podnětu citlivost.
- A (*vztahovačně*): To ale nijak neoslabuje mou pozici. Plně souhlasím s tím, že nedostatek nových podnětů se na vnímání skutečně negativně podepisuje, ale je otázka, co si pod tím kdo představí. Mně se vybaví astronauti, polární badatelé nebo třeba Natascha Kampuschová, všichni dlouhodobě pobývající ve stereotypních podmínkách. Celé dny trávené ve stále stejném prostředí, ve společnosti stejných lidí, prohlížení si týchž předmětů, provádění stále se opakujících operací. Rozmanitost možných podob objektu je takto jednotvárnou zkušeností výrazně omezená, což zákonitě postihuje šíři

a flexibilitu percepčních schémat a reprezentací. Vnímání ztrácí nejen na barvitosti, ale i na přizpůsobivosti.

B: Myslel jsem to jinak. Ve svém pojetí stavím nade vše informaci získanou z proměny. Z proměny vzhledu objektu či situace. Vstupními daty nemohou být jednotlivé sítnicové obrazy, na to jsou příliš chudé. Teprve srovnáváním obrazů v čase získáváme přidanou hodnotu (a použitelnou informaci). Způsob proměňování sítnicových obrazů skládá mnohem komplexnější informaci o podnětové situaci než obraz jeden. A ještě důležitější pro naši existenci je porozumění zákonitostem proměn sensorických dat s naším pohybem. Vy žehráte na to, jak poddeterminovaný je obraz podnětu v oku a jaká dlouhá cesta vede k výslednému vjemu. Omyl, stačí si všimnout proměn a rozumět jim v souvislosti s vlastním jednáním.

A (*trochu podrážděně*): Prostorová struktura prostředí nemůže být daty na sítnici jednoznačně vytyčená, a to ani když zohledníme pohyb. Vždy zůstanou minimálně dvě možné interpretace. Nechci nijak snižovat důležitost pohybu při vnímání, ale oproti statickému obrazu neznamena jeho zohlednění žádný kvalitativní skok. Vlastně jen dává pozorovateli možnost konfrontovat jeden sítnicový obraz sledované skutečnosti s poněkud posunutým druhým, třetím atd. obrazem. Podobně jako u binokulárního vidění, kde v mysli fúzujeme dva rozdílné obrazy, i zde je posun v promítnutých hodnotách užitečnou indicií při konstruování obrazu prostředí v mysli. Doufám, že jste v poslední větě nepřeslechl slova „indicií“, „konstruování“ a „v myslí“.

B (*naježí se a začne šlapat ještě usilovněji*): Nestačí říkat jen, že pohyb je jedním z informačních zdrojů umožňujících vnímání. Pohyb – nebo raději obecněji, změna – je pro vnímání naprosto rozhodující. Mluvil jsem už o fatálních následcích neměnnosti na funkčnost našeho poznávacího aparátu. Ještě subtilnější pro pochopení mého stanoviska jsou příklady situací, kdy je narušené chápání vztahu jednání a průvodních změn promítnutého obrazu. Popíšu 3 příklady. V prvním si aktér své jednání s proměnami stimulace vůbec nespojoval, ve druhém se tomuto vztahu teprve nově učil, ve třetím je tento vztah náhle jiný, než na jaký byl dosud v životě uvyklý. Všechny příklady dokládají, že samotná čivost smyslů ke smysluplnému, informativnímu vnímání nestačí. Jedinec může fyziologicky vidět, ale jakoby ani neviděl, neboť viděné si nedokáže složit dohromady.

A: To zní zajímavě. Poslouchám.

B: V jednom experimentu byla dvě čerstvě narozená koťata několik týdnů vězněná v košících připevněných ke kolotoči. Prvnímu z ko-

šíku ven koukala jenom hlava, druhé mělo navíc vyříznutý otvor pro nohy. To nebyla nepodstatná výhoda. Nohy se dotýkaly země, a kotě tak mohlo ovlivňovat pohyby kolotoče. Když se rozběhlo, svět před očima se změnil. Navzdory stísněným podmínkám se naučilo spojovat vlastní pohyb s proměnami zorného pole. Naopak pasivní kotě nemělo šanci vidění ovlivnit, a tak si ani nemohlo najít cestu k pochopení, jaký mají všechny ty změny před očima k jeho osobě vztah. Jednání bylo odříznuté od vnímání. Nelze se potom divit, že po vysvobození zůstalo i přes funkční zrak prakticky slepé – vráželo do předmětů, klopýtalo, při ohrožení nezvedalo tlapku na obranu.

A: Jestli ono nebude hrát roli, že ta zvířata trápili hned po jejich narození. Tedy před vytvořením si vlastní zkušenosti se zákonitostmi světa. Nebo si myslíte, že by dopad experimentu na funkčnost vnímání byl tak drtivý i ve vyšším věku?

B (*se zájmem připluje blíž*): Mluvíte nejspíš o etology tolik skloňovaném vtištění...

A: Ano. Jsem přesvědčen, že podmínkou vytvoření vyhovujících reprezentací a tedy i normálního vnímání je vystavování se bohaté, rozmanité stimulaci v senzitivním období.

B: No já nevím, stimulaci měla obě koťata stejnou. A neřekl bych, že právě rozmanitou.

A: Ale uznáte snad, že jedno z nich situace frustrovala výrazněji. A popravdě se mi moc nechce věřit, že ta aktivní kočka později viděla zcela bez následků. Napadá mě k tomu příběh jiné kočky, která byla hned po narození zavřená do válcové nádoby vymalované pouze vertikálními proužky, kde strávila řadu dní a po vysvobození byla výběrově slepá k podnětům orientovaným horizontálně a v přílehlých směrech. Jiné potíže s vnímáním přitom nezakoušela. Mé vysvětlení je takové, že po tak vyhraněné zkušenosti u ní nedošlo k aktivaci „protazení“ neuronů specializovaných na podněty v horizontální orientaci a současně neměla v mysli vybudovanou kategorii „horizontální“ se všemi prototypními i hraničními zástupci... Ale nedomlouvali jsme se, že si budeme povídat jen o lidech?

B: Ano, máte pravdu. Tak snad ke druhému příkladu. Slyšel jste, co se děje s vnímáním lidí stížených silným šedým zákalem, kterým ve vyšším věku operativně navrátí zrak? Tito pacienti sice začnou vidět, ale viděnému nedokážou porozumět. A to ani po delší době od operace. Do očí a dále do mozku jim vstupují informace, které ovšem neskládají o okolním světě úplný, koherentní obraz. Tím pádem tyto informace ani nemohou řídit jejich jednání, chybí znalosti o změnách stimulace s pohybem.

- A: No, slyšel. Ovšem volil bych úplně jiný výklad. Tak především tito pacienti i po operaci stále měli značné problémy se zrakem. Byla u nich zjištěná snížená ostrost zraku, citlivost k prostorovému a časovému kontrastu, oslabené stereovidění a periferní vidění. Celozivotně neprůchodné – u zdravého člověka opticky vodivé – prostředí zanechává následky na funkčnosti oka. To jen k vaší uvozející poznámce, že osoby uváděné ve vašich příkladech fyziologicky vidí. Ne tak docela. Ovšem co je pro mě podstatné: V nové životní situaci je pacient postavený před nutnost upravit své dosavadní představy o objektech světa, stojící převážně na sluchové a hmatové zkušenosti. Bez příslušných schémat a kategorií dopadá nově nabytá zkušenost na nepřipravenou půdu. S prodlužujícím se časem od operace se situace sice zlepšuje, ovšem kategorie zůstávají nepříliš specifické. Zvláštním problémem je rozpoznání stejného objektu v různých podobách. Jeden pacient viděl svého psa pokaždé jinak – když byl v pohybu a v klidu, ze strany a zepředu, blízko a daleko. Příčiny, proč nefunguje zázrak navrácení zraku, jsou různé.
- B (*zdá se, že přestal poslouchat*): Poslední příklad demonstruje spojení vnímání a jednání asi nejlépe. Jeden výzkumník si ozvláštnil život tím, že po dobu osmi dnů nosil brýle převracující obraz světa. Co se mu zdálo být vpravo, bylo ve skutečnosti vlevo. Viděl se zahýbat doleva, když přitom zatáčel doprava. Jeho vnímání ovšem zpočátku takto jednoduše obrácené nebylo. Spíš pokřivené – zmatečné, nestabilní, nepředvídatelné. Pokaždé jakmile pohnul hlavou, předměty před ním se podivně deformovali. Stejně tak jejich rozmístění v prostoru. Rovný pohyb se v očích zakřivoval. Stimulace přitom byla topologicky stejná jako před nasazením brýlí. Co se ale změnilo, bylo aktérovo porozumění (spíš naprostý chaos v chápání) stimulaci. Vztah mezi vlastním pohybem a průvodními senzoricnými daty byl pro něj náhle nečitelný. Nedá se říct, že by neviděl – nadále rozpoznával stromy, auta a jiné předměty –, ale nedokázal vidění používat. V prvních dnech byl plně zaměstnán konfliktem mezi očekáváním a skutečností, pohybem a změnami v zorném poli, téměř nebyl schopný souvislé aktivity. Postupně si ale na nová pravidla vztahu mezi ním a světem přivykl a začal řešit i komplexnější pohybové úlohy (např. vyhýbání se nábytku v pokoji). Problémy nastaly, jen když o prováděné činnosti začal přemýšlet. To tak trochu vystihuje rozdíl mezi našima dvěma pozicemi, nemyslíte?
- A: To byl zajímavý příklad. I když i tady bych mohl hnidopišit a ptát se, proč adaptace trvala tak dlouho, když všechny potřebné informace ke zrakem řízenému jednání byly v nezměněné míře na sítnici dostupné a ani detekce invariantů by v zrcadlových podmínkách ne-

měla být výrazně složitější? A naopak co vylučuje možnost, že adaptace byla podmíněná postupnou tvorbou nových pohybových schémat a scénářů, což naopak nějaký čas zabere? Každopádně stejně jako u předešlých dvou příkladů i zde mám pocit, že spíše než jen vztah vnímání – jednání je narušený celý vizuální systém na všech frontách fungování. Ale dobrá... řekněme, že na obecné úrovni jste mi svou představu provázanosti vnímání s jednáním, stejně jako možné následky dysfunkčnosti tohoto vztahu vysvětlil. Něco je zajímavé, něčemu se mi příliš nechce věřit. Ale stále mi chybí konkretizovaná představa, představa o senzomotorických znalostech v konkrétních situacích.

B: Pamatujete si na přistávající včelu z úvodu? Mění rychlost svého letu podle vzdálenosti podložky tak, aby udržovala vybraný optický parametr na konstantní hodnotě. Aniž by se řídila znalostí jediné fyzikální veličiny, je pro ni úspěšné zvládnutí této činnosti samozřejmostí. Naučila se totiž rozumět a řídit se informacemi z vlastního pohybu a ví, jak si celou úlohu zjednodušit. V typicky lidských aktivitách najdeme bezpočet příkladů zjednodušujících heuristik. Jednu takovou strategii používá baseballový hráč v poli při chytání odpáleného míče. Chytání baseballového míče by měla být mimořádně složitá úloha. Odpálený míč letí po balistické křivce a při odhadu místa dopadu je potřeba zohlednit rasaní odpalu, váhu míče, jeho letové vlastnosti, rotaci, odpor vzduchu, vliv větru. Jak je možné, že hráč v poli má vůbec nějakou šanci míč chytit? On opravdu s předstihem netuší, kdy a kam přesně míč dopadne, ani se o takový odhad nepokouší. Po odpalu sleduje dráhu letu míče a utíká tak, aby optická trajektorie míče byla po celou dráhu letu co možná nejjednodušší, v tomto případě přímá. Hráč neběží nejkratší cestou k místu dopadu, spíše po křivce, v souladu s promítnutou dráhou letu. Podobné úsporné strategie používají ve sportu všichni, kdo musí v pohybově náročných a současně prostorově komplexních podmínkách rychle reagovat – brankáři ve fotbalu a hokeji, pálkáři v baseballu, tenisti, stolní tenisti.

A: Proti tomu prvnímu příkladu nic nenamítám. Hráč má při chytání přeci jen dostatek času na průběžné korigování směru svého běhu, od chvíle ke chvíli může správnost svého počínání vyhodnocovat. Ve zbylých případech je ale, řekl bych, aktér odkázaný na první dojem, protože na vzniklou situaci musí zareagovat okamžitě a na zpřesňování původního odhadu nezbývá čas. Jinými slovy nezbývá čas na vzájemné ovlivňování se vnímání a jednání.

B (v euforii nejspíš něco přeslechl): To jste zmínil důležitou věc. Korekce, korigující informace, zpětná vazba o správnosti mého dosa-

vadního jednání je pro pohyb v prostředí skutečně klíčová. Řidič auta se při vjezdu do zatáčky neřídí podle momentální vzdálenosti vozidla od krajnice (nezná ji), ani podle úhlu, pod kterým je třeba volant točit (nezná poloměr zatáčky). Při řízení dbá pouze na to, aby vozidlo zůstávalo pozičně stále přibližně stejně vzhledem k silnici. Přibližování se ke krajnici nebo středové čáře je mu signálem pro úpravu kurzu vozidla. Jak prosté, že? Podobně vizuálně jednoduché je pro řidiče i brzdění. Vy byste se mě pravděpodobně pokoušel přesvědčovat, že řidič ve svých výpočtech času zbývajícího do případného střetu s překážkou vychází z odhadu momentální rychlosti svého auta a auta před ním a z jejich vzájemné vzdálenosti. Takový výpočet by ale vzhledem k omezenému času byl příliš komplikovaný, navíc k vyřešení situace stačí prostá vnímavost k jednoduchému optickému parametru, kterým je tempo zvětšování obrazu překážky. Čím rychleji se její obraz v oku zvětšuje, za tím kratší dobu dojde ke srážce. Tedy řidič se při brzdění snaží tempo udržet pod určitou hranicí, aby dobrzdil ještě včas. Podobnými postupy si člověk zjednodušuje další obecné pohybové aktivity jako chytání, odpalování, odražení, skákání, pohyb směrem k vytčenému cíli atd.

- A (*rozrušeně plave sem a tam*): Myslím, že to s tím zjednodušováním už přeháníte. K tomu poslednímu příkladu, jak jste přišel na to, že výpočet tempa nárůstu obrazu je snadný? A taky jaký je vlastně žádoucí výsledek? Vždyť výchozím bodem pro výpočet jsou oči, a ne kapota, což se může řidiči zle vymstít. Ale hlavně mi vadí, že to zjednodušování cpete všude. Snažíte se vzbudit dojem, že se člověk bez větších nesnází pohybuje prostředím, na situace, které vyžadují širší pochopení, kde by uplatnil svého ducha, prakticky nenarazí, pro případ nouze stačí mít po ruce zásobu triků. S tím pochopitelně nemohu souhlasit. Prakticky všechny vámi uváděné činnosti jsou ve skutečnosti mnohem komplexnější a jejich zvládnutí vedle bezprostřední reakce předpokládá porozumění, vytváření si vzorových schémat a reprezentací. Zmínil jste řízení auta. Ale to je nemyslitelné bez dobrého povědomí o následcích použití všech ovladačů (volant, pedál, řadící páka), bez adekvátního porozumění symbolům (značky, barvy na semaforu, blikající kontrolka) a konvenčním významům (klepání motoru, troubení druhého auta), bez schopnosti prostorového odhadu ve složité situaci (vyhnou se na úzké cestě dvě auta? stihne auto projet křižovatkou ještě na oranžovou? stihne předjet a zařadit se?). Zmínil jste příklad z tenisu. Hráči se ale nejprve musí naučit, jak mají vypadat jednotlivé údery raketou (sklon, rasance, faleš, postavení těla) a v jaké chvíli je při

hře použít, vytvořit si pohybová schémata. Bez znalostí a bez porozumění příčinnosti si práci s raketou neosvojí a tenis hrát ne naučí.

B: Můžete mít pravdu v tom, že do určité fáze sžívání se s danou pohybovou aktivitou je aktérův vklad užitečný. Jakmile je ovšem jednou zautomatizovaná, kruh sensorická informace – jednání funguje bez prostředníka.

A: To rozhodně ne! Žádná lidská činnost není tak prostá, aby ji šlo beze zbytku rozložit na automatismy. Schválně ve své polemice nechávám stranou skutečně složité prostorové úlohy propátrávání strukturovaného prostředí (hledání zavazadel na pojízdném pásu na letišti) či dokonce pohyb v něm (snaha vyznat se uvnitř budovy). Jejich řešení je bez znalostí, map a schémat zhola nemožné. Ale i když budu přemýšlet jen o jednoduchých, přímočarých aktivitách, stejně bude vámi zmiňovaný kruh bez znalostního pozadí neúplný. Řekněme, že se běžec a cyklista ocitnou ve shodné situaci, kdy se před nimi vynoří překážka. Běžci ta stejná sensorická stimulace zavelí „přeskoč“ nebo „zastav“, zatímco cyklistovi „šlápní na brzdu“ nebo „strhni říditka“. Oběma se zobrazí v očích totéž, ale každý provede něco jiného. Pro každý druh aktivity máme vytvořené jiné pohybové vzorce. Ty vycházejí ze sensorické stimulace, ovšem stejně tak z možnosti a omezení toho kterého druhu pohybu.

B: V příkladu s běžcem a cyklistou vlastně mluvíte o přímé, bezprostřední, nezpracovávané informaci, kterou poskytuje sledovaný objekt. Tato informace je obsažená v samotném vzhledu objektu (jeho tvaru, velikosti, materiálu, světelných vlastnostech) a prozrazuje pozorovateli něco o funkčnosti, o potenciálním využití objektu, jak se k němu vztahovat, co s ním podniknout. Návod na zrakové „uchození“ tedy poskytuje prostředí samo. Většina předmětů nese více než jeden takovýto význam nebo různé varianty téhož obecného významu, přičemž výsledný vjem se odvíjí od typu činnosti a způsobu interakce. Ostatně ve vašem příkladu jsou oba významy dosti podobné.

A: Pletete se. Podoba objektu může dát jen velmi obecnou představu o možnostech interakce, mnohem silnější determinantou pro osmyslení je osoba pozorovatele se svými znalostmi a preferencemi. Proč myslíte, že si na autě všimne úplně jiných detailů desetiletý kluk, opravář a autodesignér nebo na oblečení kolemjdoucích módní návrhář a homeless? Podoba vjemu je vždy ovlivněná tím, jak viděnému rozumíme, jak se podnět vztahuje k naší osobě. Přesvědčil jste mě ale v jedné věci: Věřím už tomu, že pokud v mé mysli zakoušená situace žádnou silnější odezvu nevyvolá, bude mé jednání pod přímým vlivem senzomotorických kontingencí.

- B: Vy jste mě taky o mnohém přesvědčil, i když zformuloval bych to jinak. Chování je přímo řízené stimulací, měnicími se vstupními daty ve chvílích plynulého, bezproblémového pohybu prostředím, případně ve chvílích jasně vytyčeného počátku, průběhu a cíle nějaké aktivity. V ostatních situacích (výskyt problému, nutnost komplexního řešení, neznalost činnosti) pozorovatel volí celostní přístup k podnětu; zapojuje paměť, myšlení, představitivost, případně jazyk; vytváří si obraz situace; podle potřeby upravuje vzorec chování.
- A (*usmívá se*): Nechci kverulovat – ten průsečík jste vystihl pěkně –, ale přijde mi, že spíš než o dva samostatné percepční systémy, jejichž aktivity se odvíjí od žádoucího výstupu v té které situaci (přímé jednání nebo předcházející popis problému), se jedná o systémy propojené a aktivované současně při každé (nebo alespoň většině) činnosti. Například při zatloukání hřebíku zároveň sledujeme a upravujeme pohyby ruky a kladiva a zároveň se chováme ke kladivu právě jako k nástroji určenému pro zatloukání. Oba systémy pracují v synergii.
- B (*oba se usmívají*): Hezky jste to shrnul. Vidíte, nakonec se shodneme. Vy obhajujete aktivitu mentální, já aktivitu tělesnou. Vlastně ten rozdíl není ani moc velký, zvláště když připustíme, že se jednání účastní oba prvky v součinnosti. Tak a dost řečí, pojďme aktivně na to, kvůli čemu tu jsme. Budeme hledat ostrůvek s potravou, nějak už cestu najdeme.

Shora se přiblížily experimentátorovy ruce, které vytáhly oba potkany z nádoby ven, a ozvalo se klení „Co že jsou ty potvory dneska tak líný?“

Poděkování: Příspěvek byl zpracovaný jako součást výzkumného záměru PsÚ AV ČR (reg. číslo AVOZ70250504) a podpořený grantem GA ČR 406/07/1676.

Základní literatura k tématu:

- [1] Ajvaz, M., Havel, I.M., Mitášová, M.: *Prostor a jeho člověk*. Praha: Vesmír 2006.
- [2] Gibson, J.J.: *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin 1979.
- [3] Gregory, R.L.: *Eye and Brain: The Psychology of Seeing*. Oxford: Oxford University Press 1998.
- [4] Harris, L.R., Jenkin, M.: *Vision and Action*. Cambridge: Cambridge University Press 1998.
- [5] Milner, A.D., Goodale, M.A.: *The Visual Brain in Action*. Oxford: Oxford University Press 1995.
- [6] Noë, A.: *Action in Perception*. Cambridge, MA: MIT Press 2004.
- [7] Norman, J.: Two visual systems and two theories of perception: An attempt to reconcile the constructivist and ecological approaches. *Behavioral and Brain Sciences* **25** (2002) 73-144.
- [8] O'Regan, K., Noë, A.: A sensorimotor approach to vision and visual consciousness. *Behavioral and Brain Sciences* **24** (2001) 939-973.
- [9] Rock, I.: *The Logic of Perception*. Cambridge, MA: MIT Press 1983.

Specifikace výpočetního modelu mysli vědomého vtěleného agenta

Jiří Wiedermann¹

Hjakudzó svolal žáky a postavil před ně džbán: Kdo mi řekne, že to je džbán, aniž ho pojmenoval? Isan předstoupil a kopl do džbánu, až ho převrhl.

Abstrakt. Nastíníme jednoduchou, ale přesto kognitivně účinnou architekturu inteligentního agenta. Model využívá dvou komplementárních vnitřních modelů světa: jeden pro „syntax“ poznaného světa a druhý pro jeho sémantiku. Tyto modely řeší problém porozumění konceptům a podporují algoritmické procesy, jejichž efekty se pro pozorovatele jeví jako projevy vyšších kognitivních funkcí, jakými jsou imitační učení, rozvoj komunikace, řeči, myšlení a vědomí.

1 Úvod

Citát z úvodu z práce předního českého japanologa Miroslava Nováka, *Každý sám svým pánembohem* [15] naznačuje, že mistři filosofie zen znali klíč k řešení problému, který se v umělé inteligenci skrývá pod názvem „*problém ukotvení symbolů*“. Tento problém se týká otázky, jak slova získávají svůj význam a co to vlastně ten význam je [9]. Umělá inteligence se tímto problémem začala zabývat v osmdesátých letech dvacátého století, kdy se tento problém manifestoval v podobě „bajky“ o tzv. čínské komnatě, kterou „vymyslel“ filosof J. Searle [22]. Tato bajka ilustrovala problém, jestli počítač může rozumět slovům (tj. symbolům), pomocí kterých komunikuje s lidmi. Searle upozornil na skutečnost, že počítač nemůže se slovními symboly dělat nic jiného, než je podle nějakých pravidel transformovat na jiné symboly (v konečném důsledku na posloupnosti nul a jedniček), kterých sémantiku počítač také nemůže znát. Takže, alespoň podle Searla, „*tudy cesta nevede*“. Tato bajka rozvířila snad do dnes neutuchající diskusi a vedla k rozvoji teorií, které tvrdily, že proto, aby uměle-inteligentní systémy „rozuměly“ tomu, co dělají, potřebují mít tělo. Takovým systémům se někdy říká „vtělení agentů“. Teprve tělo, obvykle v interakci s vnějším světem, dává těmto systémům prostředek, jak porozumět svým vlastním akcím i akcím ostatních vtělených agentů, a jak s nimi komunikovat. To ostatně využil i žák Isan z našeho citátu: nemít těla a nebýt džbánem, nemohl by kopt do džbánu, ten by se nepřevrátil a nevydal by přitom charakteristické zvuky a nevyvolal by v přihlížejících soubor příslušných

¹ Ústav informatiky AV ČR, v.v.i., Pod Vodárenskou věží 2, 182 07 Praha 8, Česká republika, e-mail: jiri.wiedermann@cs.cas.cz

paměťových a smyslových vjemů, které jsou právě oním hledaným významem slova „džbán“. Poeticky řečeno, Isan tedy jednoduše u přihlížejících „zahrál na stejnou strunu“, kterou by v nich rozezvučelo nevysslovené slovo „džbán“.

Vraťme se ještě na chvíli k Searlovu počítači, který proto, aby komunikoval s lidmi „inteligentně“, musel zjevně mít nějakou informaci o tom, jak vypadá vnější svět, jaké asi jsou schopnosti a pocity lidí v různých situacích, atd. Zkrátka, musel mít jakýsi vnitřní model vnějšího světa, jakkoliv reprezentovaný. Tento model pak tvořil součást pravidel, pomocí kterých by Searlův počítač komunikoval s lidmi,

Myšlenka, že netriviální inteligentní systémy by měly využívat nějakou formu vnitřního modelu světa, sahá do prvopočátků umělé inteligence. Nicméně, pokusy o řízení chování pomocí formálních pravidel pracujících nad symbolickými modely světa selhaly. Následkem toho se hlavní proud výzkumu obrátil v poslední dekádě minulého století směrem k biologicky motivovaným modelům založeným na přímém napodobování chování živých tvorů, většinou hmyzu. Tento přístup zdůrazňoval nutnost ztělesnění a situovanosti pro využití chování jednoduchých robotů řízených sensorickými podněty (viz např. [4]). Toto paradigma fungovalo dobře zejména ve spojení s tzv. subsumpční architekturou, která využívala na sebe navazující vrstvy stále složitějšího chování, které bylo programováno na míru požadovaným stále náročnějším úkolům (viz např. [17]). Bohužel, po počátečních úspěších, převážně při konstrukci různých robotů reagujících na přímé podněty se narazilo na meze takového přístupu. Zjevné to bylo zejména v humanoidní robotice, kde se další postup směrem k vyšším kognitivním funkcím zdál nemožný bez zásadních inovací kognitivní architektury příslušných agentů. Možná, že zde je třeba hledat důvody útlumu konstrukčních aktivit v oblasti humanoidní robotiky, který se ale na druhé straně zdá být kompenzován nárůstem počtu teoretických prací v přílehlých oblastech, kladoucích si za cíl umělé vědomí (viz např. [10]).

V současné době převažuje názor, že pro prolomení dříve zmíněné bariéry, na kterou narazila robotika přímými podněty řízených robotů, a k otevření cesty k vyšším mozkovým funkcím potřebujeme automatické mechanismy rozšiřující dříve získané znalosti a dovednosti robotů [23]. Takové mechanismy mohou využívat vhodné vnitřní modely světa. Dnes převažující trendy dávají přednost tzv. sub-symbolickým modelům (především neuronovým sítím) před tzv. symbolickými modely, u kterých je přímočarý vztah mezi objekty a jejich reprezentací. Přehled současného stavu a diskusi o vnitřních modelech světa lze nalézt např. v pracích [5] a [11].

V práci [11] její autoři Holland a Goodman argumentují ve prospěch interního modelu světa sestávajícího ze dvou oddělených, avšak spolupracujících částí: modelu agenta a modelu jeho okolí. Nedávno byl v oblasti teoretické informatiky publikován podobný model [3] sloužící k úvahám o definici vědomí. V těchto a podobných pracích se jejich autoři domnívají, že klíč k pochopení vědomí je ukryt v definici a ve funkci výše zmíněného dvousložkového modelu. Cruse [5] přichází k podobnému závěru použitím vnitřního modelu, který zachycuje pouze agentovo vlastní tělo.

Tato práce navazuje na práce autorů zmiňovaných v předchozím odstavci. Metodologickým východiskem pro náš přístup bude postup používaný v softwarovém inženýrství při návrhu velkých systémů. Začneme nástínem architektury kognitivního systému a uvedeme neformální funkční specifikace jednotlivých jeho modulů. To znamená, že definujeme typ dat a jejich tok mezi jednotlivými moduly a také úkol jednotlivých modulů z hlediska zpracování příslušných dat. Dále již budeme argumentovat ve prospěch uvedeného modelu — jaké jsou důvody domnívat se, že model podporuje realizaci procesů, o kterých se lze oprávněně domnívat, že odpovídají vyšším kognitivním schopnostem, jakými jsou např. imitační učení a rozvoj komunikace, řeči, myšlení a vědomí.

Naše kognitivní architektura bude vycházet z myšlenky dvou spolupracujících vnitřních modelů. První model bude tvořen tzv. zrcadlovou neuronovou sítí, která se bude učit často se opakující „percepčně-behaviorální“ jednotky. Tyto jednotky jsou reprezentovány v modelu pomocí tzv. multimodální informace, která je fúzí senzorických a motorických informací vztahujících se k „jednotce“ situace. Návrh zrcadlové neuronové sítě, která je odpovědná za agentovu situovanost v jeho prostředí, byl inspirován zjištěnými vlastnostmi nedávno objevených biologických zrcadlových neuronů v mozku primátů [19], [20] a také nejnovějšími neurofysiologickými objevy [7], které naznačují, že zrcadlové neurony skutečně reprezentují kognitivní mechanismus pro porozumění akcím, úmyslům a emocím, které jsou vybuzeny senzorickými stimuly. Zrcadlová neuronová síť reprezentuje v jistém smyslu jak agenta, tak i jeho okolí; zachycuje současně syntaxi i sémantiku korektního chování. V odpovídající multimodální informaci je svět reprezentován senzorickými vstupy a agentovy akce jsou zachyceny v odpovídajících motorických instrukcích a jeho „pocitech“ danými zpětnovazební informací od jeho vnitřních sensorů. Takže, na jedné straně, zrcadlová síť zachycuje podobnou informaci jako Cruseho vnitřní model (což je vlastně také neuronová síť) agentova těla anebo model agenta v Hollandově a Goodmanově modelu. Na druhé straně, protože v zrcadlové síti jsou také

prvky environmentální informace zprostředkované agentovými senzory, tato síť jistým fragmentovaným způsobem také reprezentuje okolí (na sub-symbolické úrovni) podobně jako druhá část Hollanova a Goodmanova modelu.

Druhý vnitřní model světa je tvořen agentovou řídicí jednotkou. Tato jednotka kontinuálně zpracovává multimodální informaci dodávanou zrcadlovou sítí. Úkolem řídicí jednotky je dolovat znalosti z toku multimodálních informací. V řídicí jednotce jsou znalosti reprezentovány pomocí rekurentní sítě konceptů. Základní koncepty jsou tvořeny jednotkami multimodálních informací. Dále řídicí jednotka také automaticky odvozuje koncepty, které neodpovídají žádné multimodální informaci, nýbrž reprezentují znalost odvozenou, abstrahovanou ze základních konceptů. Řídicí jednotka odhaluje pomocí statistických mechanismů často se vyskytující vzory v toku základních konceptů a na základě těchto vzorů formuje abstraktnější koncepty a učí se jejich časové či prostorové vztahy. To znamená, že příslušná síť konceptů se vlastně učí různým vzorům chování. Vycházejíc z multimodální informace o současné situovanosti, řídicí jednotka určí následující akci agenta. Také řídicí jednotka je implementována pomocí rekurentní neuronové sítě. Je zřejmé, že řídicí jednotka zachycuje dynamické aspekty agentovy interakce s jeho okolím a jako taková nemá protipól v modelech autorů zmiňovaných předtím.

Náš model umožní věrohodné vysvětlení výpočetních mechanismů stojících za jevy, které se ve svých důsledcích podobají vyšší mozkové činnosti, včetně vědomí. V našem modelu je výpočetní vědomí chápáno jako poslední fáze posloupnosti postupně stále více kognitivně náročnějších schopností systému rozvíjejících se ve stále více stimulujícím prostředí. Odpovídající posloupnost začíná na úrovni schopnosti učení pomocí imitace, pokračuje přes schopnost naučit se a porozumět řeči těla, posunkům a artikulované komunikaci mezi příslušníky stejného druhu, a dále vede přes schopnost mluvení sama k sobě k myšlení. Ve finále tento vývoj vede do stavu, ve kterém jsou kognitivní entity schopné popsat ve vyšším jazyku libovolnou minulou, přítomnou anebo očekávanou událost a přemýšlet o nich („*generovat na míru šité příběhy*“, jak to nazývá Blum ve své práci [3]). Také v našem modelu se tento stav považuje za příznak vědomí. To ostatně dobře odpovídá Minského poznámce o tom, že „*vědomí je velký kufřík*“, obsahující mnoho různých mentálních schopností [14].

Myšlenka, že zrcadlové neurony jsou klíčem k imitačnímu učení a že hrají důležitou roli při rozvoji přirozeného jazyka, počala klíčit začátkem tohoto století (viz. např. [1], [8], [13], [18], [19]). Jeden z prvních výpočetních modelů, založených na zrcadlových neuronech,

byl publikován v práci [27]. V nynější práci je rozpracována myšlenka chápání zrcadlových neuronů jako vnitřního „syntaktického“ modelu světa, který společně s vnitřním sémantickým modelem světa skýtá rámec pro řešení problému ukotvení symbolů a pro rozvoj výpočetního vědomí; dále je zde prezentován nástin příslušných kognitivních algoritmů včetně definice výpočetního vědomí. Práce představuje rozpracování autorových myšlenek, prezentovaných v [26], [27], [28] a [29]. Naše výsledky potvrzují konstruktivním způsobem intuici dřívějších badatelů, že totiž máni těla a vnitřní modely světa představují základ pro rozvoj vyšších mentálních funkcí, včetně vědomí.

Struktura článku je následující: ve 2. části představíme podrobněji náš model. Ve 3. části popíšeme jeho fungování vedoucí ke vzniku výpočetního vědomí.

2 Model

Struktura modelu je znázorněna na obr. 1. Model se skládá ze 4 hlavních součástí: senzomotorické jednotky, syntaktického modelu světa reprezentovaného zrcadlovou neuronovou sítí, sémantického modelu světa (řídící jednotky) a těla. Tok (digitálních) dat mezi jednotlivými moduly je zobrazen šipkami.

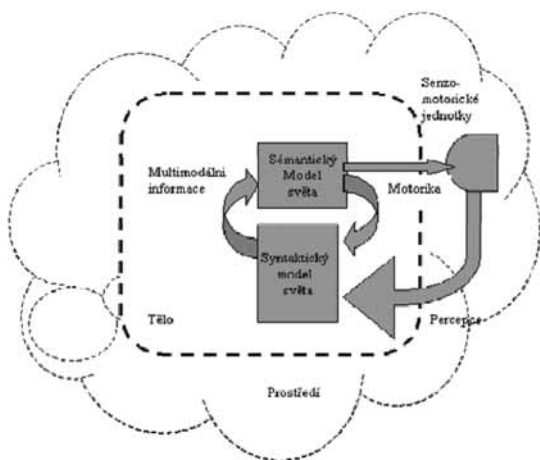
Senzomotorické jednotky dostávají od řídicí jednotky *motorické instrukce*. Nejsou to pouze instrukce pro agentovy lokomoční orgány, ale také instrukce určené senzoričtým orgánům sloužící k nastavení jejich parametrů: zacílení na určitý cíl, zaostření apod. Kopie těchto instrukcí jsou současně nasměrovány do zrcadlové sítě. Senzomotorické jednotky posílají do zrcadlové sítě dva druhy dat.

Prvním druhem jsou tzv. *extracepční data* nesoucí informaci od senzoričtých jednotek, které zkoumají agentovo okolí. V tomto případě reagují senzory na fyzikální vstupy (elektromagnetické vlnění, zvuky, tlak, teplota, atd.) a transformují je do digitální podoby. Druhý typ dat jsou tzv. *propriocepční data* pocházející od senzomotorických jednotek agenta rozmístěných uvnitř jeho těla. U člověka je příkladem takové informace vnitřní teplota, krevní tlak, svalový tonus, apod.

Další součástí modelu je *zrcadlová síť*. Je to síť umělých neuronů, které modelují chování reálných zrcadlových neuronů. V každé jednotce této sítě (skládající se z několika neuronů) se setkávají extracepční a propriocepční data od senzomotorických jednotek s motorickými instrukcemi z řídicí jednotky. Jednotka tvoří jejich konjunkci, která se nazývá *multimodální informace*. Zrcadlová síť plní tři hlavní úkoly:

- *Učení*: pamatuje si často se vyskytující multimodální informace;
- *Identifikace*: k dané vstupní multimodální informaci síť najde „nejpodobnější“ multimodální informaci zapamatovanou v síti;

- *Asociativita*: v případě, že do sítě vstoupí neúplná či poškozená multimodální informace, síť doplní její chybějící či poškozené části.



Obrázek 1. Vtělený kognitivní agent se dvěma vnitřními modely světa

Aby mohla síť tímto způsobem pracovat, musíme zařídit, aby si pamatovala pouze konečné množství „důležitých“ informací. To lze dosáhnout vhodnou parametrizací percepčních dat a motorických instrukcí pomocí konečné množiny hodnot parametrů. Jinou možností je využití fuzzy přístupů, kterých efekt je hrubá klasifikace dat do konečného počtu shluků podobných multimodálních dat. Pro správnou práci asociativního mechanismu zrcadlové sítě je nezbytné, aby multimodální informace obsahovala co nejvíce redundantních údajů, které umožní zrekonstruovat celou informaci v případě, že některé její části chybějí. Proto je důležité, aby agent byl vybaven dostatečným množstvím senzorů, poskytujících k jedné „události“ informace různého typu. Speciálně budeme předpokládat, že motorické údaje samotné stačí pro rekonstrukci zbytku multimodální informace. Tento předpoklad budeme potřebovat pro práci algoritmu myšlení.

Každá jednotka zrcadlové sítě se specializuje na naučení a poté i rozeznání specifické multimodální informace, která odpovídá „*percepčně-behaviorální jednotce*“. Učení se děje průběžně v případě, kdy do sítě vstupují „nepoškozené“ kompletní multimodální informace. Taková situace se nazývá *standardním učícím režimem*. Učení postupuje pomocí tzv. hebbovských principů, tj. posilováním vah neuronů

kteře reprezentují danou multimodální informaci vždy, když se taková informace rozezná.

Pokud funguje zrcadlová síť naznačeným způsobem, tak jejím výstupem je proud kompletních multimodálních informací bez ohledu na to, jestli informace vstupovala původně do sítě poškozena anebo neúplná. Tento proud teče do řídicí jednotky; zde se jednotky multimodální informace nazývají *koncepty*. Úkolem řídicí jednotky je dolovat znalosti ze vstupujícího proudu konceptů a aktivovat jiné koncepty, zapamatované v řídicí jednotce. Motorická část aktivovaných konceptů je pak zaslána senzomotorickým jednotkám a její kopie do zrcadlové sítě. Řídicí jednotka tedy určuje další akci agenta.

V řídicí jednotce existují koncepty odpovídající každému výskytu multimodální informace vystupující ze zrcadlové sítě. Navíc se v řídicí jednotce formují nové tzv. *abstraktní koncepty*, které bezprostředně neodpovídají žádné konkrétní multimodální informaci. Koncepty jsou v řídicí jednotce propojeny přes vazby zvané *asociace*, které jsou ohodnoceny *vahami* různé velikosti. Koncepty společně s asociacemi a jejich vahami tvoří agentovu paměť.

Pravidla pro formování nových konceptů, vznik asociací a velikost jejich vah jsou založena na principech, kterých původ sahá až k anglickému filozofovi 18. století, D. Humovi [12]:

- *Současnost*: dva koncepty budou asociovány (anebo, pokud již jsou asociovány, tak váha asociace vzroste), pokud se často vyskytují současně; souběžně vznikne nový koncept, který je sjednocením obou konceptů. Tento nový koncept je *konkretizací* každého z původních dvou konceptů,
- *Souslednost*: dva koncepty budou asociovány (anebo, pokud již jsou, tak váha asociace vzroste), pokud se často vyskytují jeden po druhém.
- *Podobnost*: koncept bude asociován s jiným konceptem, pokud jsou si podobny; míra podobnosti je definována velikostí překryvu příslušné multimodální informace.
- *Abstrakce*: společná část dvou konceptů tvoří koncept, který je abstrakcí obou konceptů; tento koncept se přidá k množině původních konceptů.

Kontrolní jednotka pracuje dle následujících pravidel. V každém okamžiku některé její koncepty jsou v aktivním stavu. Tyto koncepty reprezentují současný „*mentální stav*“ agenta. Vstupující multimodální informace aktivuje další koncepty. Nový mentální stav se počítá pomocí asociací, které jsou mezi aktivními a ostatními koncepty. Aktivní koncepty excitují koncepty, se kterými jsou asociovány silou

úměrnou vahám jednotlivých asociací. Tato excitace samotná mírně posiluje váhy příslušných asociací. Váhy ostatních asociací jsou naopak mírně sníženy — to modeluje proces zapomínání. V množině excitovaných konceptů se vyberou nejvíce excitované koncepty a tyto jsou aktivovány; ostatní koncepty jsou deaktivovány. Množina nově aktivovaných konceptů určuje nový mentální stav a také další motorické akce agenta. Na množinu aktivních konceptů lze nahlížet jako na *krátkodobou (operační) paměť*. Množina všech konceptů, asociací a jejich vah odpovídá *dlouhodobé paměti agenta*.

Je zřejmé, že mechanismus výpočtu nového mentálního stavu připomíná mechanismus výpočtu konečného automatu a že jej také lze implementovat podobně jako model zrcadlových neuronů — tj. pomocí neuronové sítě.

Na základě předchozích principů je řídicí jednotka schopna realizovat jednoduché kognitivní úkoly, jako např. učení současně se vyskytujících konceptů (pomocí současnosti), jejich posloupnost (tzv. jednoduché podmiňování pomocí souslednosti), podobné chování a počítat jejich abstrakce. Mechanismus je dokonce schopen realizovat *pavlovské reflexy* (viz např. [25], str. 217), kdy je agent podmíněn produkovat jisté chování v odpovědi na zdánlivě nesouvisející podnět.

Pokud chceme pokročit směrem k realizaci náročnějších kognitivních úkolů, musíme zavést speciální koncepty zvané *afekty*. Tzv. základní afekty tvoří podmnožina konceptů aktivovaných prostřednictvím senzorů. Existují dva druhy afektů: *pozitivní*, odpovídající příjemným či žádoucím pocitům, a *negativní*, odpovídající negativním pocitům. Asociace vycházející z afektů nesou „znaménko“ afektů a excitace přes takové asociace modulují excitaci cílového konceptu — mohou ji posílit anebo utlumí. Pomocí afektů lze simulovat *učení pomocí odměn anebo trestů*, které následují buď bezprostředně, anebo až s jistým zpožděním po vykonání akce. Zdá se, že schopnost pavlovských reflexů a učení pomocí odměny a trestu představují kognitivní minimum, které musí agent splnit, pokud aspiruje na netriviální chování.

Během agentovy interakce se stimulujícím prostředím se koncepty v rámci řídicí jednotky začnou samoorganizovat do *klastrů* definovaných na základě podobnosti. Jádra těchto klastrů jsou tvořena koncepty, které jsou největší společnou abstrakcí ostatních konceptů v klastru a stále si zachovávají jistou netriviální míru podobnosti s ostatními koncepty v klastru. Pomocí časové souslednosti se jádra klastrů řetěží do tzv. *zvyků*. Jakmile se zvyky dostatečně upevní, chování agentů se jimi začne řídit. Takové chování se ve většině případů odvíjí bez jakéhokoliv „mentálního“ úsilí. Pouze v situacích, kdy se zvyk

v nějakém kontextu „kříží“ s jiným zvykem, si řídicí jednotka vyžádá další informace (např. v tzv. off-line režimu od zrcadlové sítě, anebo v on-line režimu od senzomotorických orgánů — viz v dalším), které upřesní kontext a umožní výběr odpovídajícího chování. Více detailů o práci řídicí jednotky a jejich kognitivních schopnostech viz v práci [26], [28] a tam uvedených odkazech na literaturu.

Poslední složkou našeho modelu je *tělo*. Tělo je nositelem senzomotorických orgánů agenta a současně tvoří ochrannou schránku všech jeho dalších součástí.

Nyní se vraťme k otázce vnitřních modelů. Na zrcadlovou síť můžeme zjevně pohlížet jako na syntaktický model světa. V tomto modelu je svět reprezentován tak, jak je poznán prostřednictvím agentových smyslů a jeho motorických akcí. Tento model můžeme nazvat i *senzomotorickým modelem*, protože popisuje „syntaxi“ agentova světa. V zrcadlové síti jsou uloženy ty kombinace exterocepčních a propiocepčních vstupů a motorických instrukcí, které se „hodí k sobě“. Všimněme si, že součástí multimodální informace jsou vždy i propiocepční informace, které se samozřejmě bezprostředně a výlučně týkají agenta samotného. To znamená, že v zrcadlové síti jsou přítomny i prvky modelu samotného agenta.

Na druhé straně, řídicí jednotka je specifickým modelem světa, který zahrnuje „sémantiku“ světa. Asociace a koncepty uložené v tomto modelu zřejmě odpovídají relacím a jevům ve skutečném světě tak, jak je agent vnímal pomocí svých sensorů a akcí. To vše vlastně reprezentuje *vnitřní dynamický model světa*. Na tento model lze také nahlížet jako skladiště „vzorců“ chování, která jsou smysluplná v daných situacích.

V další části popíšeme, jak interakce obou částí modelu vede k složitějšímu chování.

3 Vyšší kognitivní funkce

Nejprve podáme nástin mechanismu imitačního učení, které představuje základ vyšších mentálních schopností (viz např. [1], [11]). Představme si následující situaci: agent *A* pozoruje agenta *B*, který koná jistý dobře rozlišitelný specifický úkol (např. napřahuje ruku po banánu). Pokud *A* má ve svém repertoáru percepčně-behaviorálních jednotek chování situaci, která se dobře shoduje s pozorováním zprostředkovaným senzory agenta *A*, tak jeho zrcadlová síť tuto jednotku identifikuje pomocí asociačního mechanismu, pouze na základě senzorní informace. Současně, protože v odpovídající percepční informaci chybí propiocepční složka, asociativní mechanismus ji doplní (na základě podobnosti s minulou agentovou zkušeností)

a navíc přidá i jednobitový příznak znamenajícím „*toto není má vlastní akce*“. Asociativní mechanismus také doplní i chybějící motorickou informaci, tj. „ukotví“ vnímané vjemy do tělesných akcí (ale tyto akce, samozřejmě, nerealizuje). Takto zkompletovaná multimodální informace postoupí do řídicí jednotky, kde je adekvátním způsobem zpracovaná. Řídicí jednotka agenta *A* má v tuto chvíli k dispozici informaci o tom, v jaké je *B* situaci, a z ní může odvodit (jako by sám *A* byl v situaci, ve které je *B*), co pravděpodobně udělá *B*. Toto „předjímání“ je realizováno jednoduše sledováním asociací, které začínají v současném mentálním stavu agenta *A*. Agent *A* dokonce má k dispozici informaci o „pocitech“ agenta *B* — jednoduše jsou to propriocepční informace (resp. příslušné afekty) agenta *A*, které doplnil asociativní mechanismus, a samozřejmě také zmíněný příznak, který v tomto kontextu znamená „*toto nejsou mé pocity*“. Toto můžeme považovat za model „*vcítění*“ (empatie) v našem zjednodušeném modelu. Navíc, pokud nadáme naše agenty schopností zapamatovat si krátké nedávně pozorované posloupnosti akcí, tak *A* může zopakovat akce, které pozoruje u agenta *B*. Za strany agenta *A* to ovšem není nic jiného než *imitace* chování agenta *B*.

Stejný mechanismus pomáhá formaci detailnějšího modelu agenta samotného — tzv. koncept *self*. Funguje to takto. Pozorování aktivit agenta stejného druhu jiným agentem umožňuje pozorovateli „doplňovat si“ mezery v jeho vlastním dynamickém modelu světa, protože již od samého počátku má pozorovatel k dispozici informaci o tom, „*jaké to je, vnímat svou vlastní motoriku*“, a teď ji doplňuje tím, „*jaké to je, když vnímám tutéž motoriku u jiného*“. V této fázi jsme již u *primitivní komunikace* prováděné pomocí *gest*, anebo obecněji, pomocí *řeči těla*. Naznačením nějaké akce agent vysílá vizuální informaci, která je doplněna pozorovatelovou zrcadlovou sítí na úplnou multimodální informaci. To znamená, že pomocí jediného gesta může být sdělena komplexní informace. V tomto případě tedy gesto zastupuje prvek jakéhosi (proto)jazyka vyšší úrovně. Mimochodem, zde mohou do komunikace vstoupit i *výpočetní emoce* jako jedna ze složek komunikace. Jejich účelem je modulovat agentovo chování. Samozřejmě, že za tím účelem musí být agenti vhodným způsobem vybavení (např. specifickou mimikou, možností změny barvy, apod.). Jakmile má agent schopnost artikulace, může doplnit gesta, a později je dokonce nahradit, *artikulovanými zvuky*. To lze chápat jako zrození mluvené řeči. Někdy v této době začíná proces stratifikace abstraktních konceptů od vtělených a agent začíná „rozumět“ gestům (tj. jazyku těla), což jsou vlastně ztělesněné abstraktní koncepty. Je dobré si uvědomit, že agenti rozumějí gestům a řeči prostřednictvím „*vcítění*“

se“ do takové komunikace, v termínech ukotvení obsahu komunikace ve stejné senzomotorice [8][9] a ve složitějších případech, ve stejných zvucích (viz např. [9]). Dále je třeba si uvědomit, že agent, který by alespoň ve fázi poznávání světa a učení porozumění svým vlastním akcím a vjemům neměl tělo, by se nemohl naučit rozumět komunikaci (viz např. [16]).

Pokud se vrátíme k procesu stratifikace abstraktních a vtělených konceptů, tak na tyto dvě třídy konceptů lze nahlížet jako na koncepty na symbolické a sub-symbolické úrovni. Tento náhled dává odpověď na často zmiňovaný problém, jestli mysl pracuje s jednou či druhou třídou konceptů: náš model pracuje s oběma třídami a přechází plynule od jedné ke druhé. Sub-symbolická úroveň vtělených konceptů je nutná pro porozumění abstraktním konceptům. Jazyk je potom vlastně nadstavbou nad vtělenými koncepty. Přejít od gest a případně řeči těla k artikulované řeči neznamená pouze to, že se gesta „naváží“ na příslušné zvuky, ale především na motoriku mluvidel. To dále umožní „samomluvu“ (hovoření sama k sobě) a později umožní přechod k myšlení (viz dále).

Vycházejíc ze struktury a funkčnosti řídicí jednotky a zrcadlové sítě, náš model realizuje řešení problému ukotvení symbolů v podobném duchu, jaký byl naznačen (ale nedořešen do takové úrovně, jako v našem modelu) v práci [23]. Výše naznačená realizace vyšších kognitivních funkcí je v dobrém souladu s tzv. teorií intencionálního souznění, která vychází z předpokladu, že sdílené vzorce neurální aktivity a doprovodní vtělené simulace tvoří biologický základ pro porozumění záměrům jiných agentů [7].

Agent schopný komunikace v našem smyslu je jenom krůček od myšlení. V našem modelu je *myšlení realizováno jako mluvení se sebou samým*. Agent mluvící sám se sebou spustí mechanismus rozlišující mezi externími stimuly („*poslouchám mluvenou řeč*“) a interními („*jsem to já, kdo mluví*“). Zde je základ *uvědomění* v našem modelu. Další malou modifikací (z hlediska inteligentního designéra agenta) můžeme dosáhnout, že agent do samomluvy ani nemusí zapojovat svá mluvidla. V tomto případě příslušné instrukce se nedostanou k tomuto orgánu, pouze jsou přímo usměrněny do zrcadlové sítě (viz obr. 1). Zde vybudí stejnou multimodální informaci, jako v případě, kdyby agent slyšel příslušné artikulované zvuky anebo vnímal řeč svého těla prostřednictvím propriocepce. (Zde využíváme naši poznámku o tom, že motorická část multimodální informace podmiňuje její zbytek). Zřejmě během myšlení agent „odpojí“ jakoukoliv interakci s vnějším světem (tj. percepci a motorické akce). Agentu v *režimu myšlení* znázorňují tmavé šipky na obr. 1 představující cyklus

z kontrolní jednotky do zrcadlové sítě a zpět. Všimněme si, že z hlediska interního mechanismu pracuje agent v režimu myšlení podobně jako ve standardním režimu učení. Rozdíl je v tom, že v posledně zmíněném případě agent pracuje s reálnou percepcí a vykonává všechny motorické instrukce, kdežto v předchozím případě běží stejné „mentální“ procesy. Tyto procesy vycházejí z virtuálních, nikoliv skutečných, dat uložených v zrcadlové síti; příslušné motorické instrukce se nerealizují. Řečeno počítačovou terminologií, v režimu myšlení agent pracuje *off-line*, kdežto ve standardním režimu pracuje *on-line*. Ještě si všimněme, že pokud má agent schopnost odpojit se od reality v režimu myšlení, pak agent rozlišuje mezi myšlením a realitou. To se považuje za základ vědomí [21].

V našem modelu budeme definovat výpočetní vědomí v duchu Minského ideje, že „*vědomí je velký kufř*“ obsahující mnoho různých mentálních schopností. Prologem k vědomí je komunikace a myšlení. Následující „definice“ výpočetního vědomí předpokládá, že agent je schopen komunikovat v abstraktním vyšším jazyku. Vyšší jazyk je „abstraktní“ jazyk používající slovní výraz či gesto k označení relativně složité akce (odpovídající posloupnosti mentálních stavů) anebo abstraktnímu konceptu. Úroveň jazyka je tím vyšší, čím bohatší je jazyk, tj. čím větší a abstraktnější je množina věcí a událostí, o kterých lze v jazyce komunikovat. Budeme říkat, že *agenti mají vědomí*, pokud jejich jazykové schopnosti dosahují takové úrovně, že agenti jsou schopni fabulovat na dané téma. Přesněji, agenti jsou schopni:

- Mluvit, přemýšlet a vysvětlit z hlediska 1. anebo 3. osoby minulé, přítomné anebo očekávané zážitky, pocity, záměry a pozorované jevy;
- Napodobit pozorované akce jiných agentů, verbálně je popsat, a naopak realizovat akce na základě jejich verbálního popisu ve vyšším jazyku;
- Rozšiřovat svůj jazyk vyšší úrovně o nová slova, učit se novému jazyku.

Zdá se tedy, že stav vědomí nelze dosáhnout bez toho, že by agent měl k dispozici vnitřní model světa doplněný znalostmi o tom, jak svět funguje a jak funguje agent sám (a jemu podobní) v tomto světě. To je nemyslitelné bez toho, aby agent měl schopnost učení. Nutným požadavkem pro vznik vědomí je sociální interakce agentů ve vyšším jazyce se stejnou či podobnou sémantikou. Zřejmě vědomí není vlastnost, kterou entita buď má, anebo nemá. Tuto vlastnost může agent mít v různé míře. Např. z předchozích požadavků vyplývá, že vědomý agent by měl být schopen rozeznat sám sebe v zrcadle; ještě „vědo-

mější“ agent by měl být schopen rozumět pohádkám, ale „vědět“, že to jsou smyšlené historiky. Ještě více vědomý agent by měl umět lhát (nikoliv např. pomocí mimikrů, ale ve vyšším jazyku!) a být si toho vědom, atd. A jak zjistíme, jestli vědomý agent „rozumí“ tomu, co říká? Zřejmě nejjednodušší bude, zeptat se ho. Schopnost odpovídat na takové otázky je vlastně jádrem naší definice vědomí.

Právě popsaná definice vědomí je vlastně testem, který může aplikovat entita, která si myslí, že má vědomí, na jinou entitu, aby rozhodla, jestli jiná entita má také vědomí. Dle této definice postupujeme i my, lidé. Nicméně, uvědomme si, že v článku jsme dokázali něco více, než že jsme dospěli k definici výpočetního vědomí. Zdůvodnili jsme, že kognitivní agent s navrhovanou architekturou a právě popsanou funkčností jednotlivých jejích modulů v principu splňuje všechny předpoklady pro to, aby byl vědomý. Jestli se vědomí rozvine je pak otázkou agentova „správného ztělesnění“, vhodných technických parametrů jeho orgánů (paměťové kapacity a efektivity, operační rychlosti, vlastností senzomotorických jednotek, atd.), a také, samozřejmě, je to otázka správné výchovy agenta. Tento přístup připomíná poměry v teorii výpočtů: každý správně navržený počítač (Turingův stroj, anebo osobní počítač, řekněme), může být v principu univerzálním počítačem; nicméně, proto, aby fungoval, musí být patřičným způsobem sestaven a naprogramován. Stejně závěry platí i pro náš model vzhledem k myšlení a vědomí. Podobná myšlenka, totiž, že „zjistit, jestli nějaký mechanismus může podporovat rozvoj vědomí v nějakém organizmu, lze pouze analýzou tohoto mechanismu“ byla zmíněna např. v práci [1], avšak příslušný mechanismus nebyl naznačen.

Věříme, že náš nástin modelu vědomého kognitivního agenta představuje první krok směrem k teorii kognitivních systémů, o kterých nebude třeba rozhodovat pouze pomocí testů, jestli mají nějaké kognitivní schopnosti, ale na základě zkoumání jejich architektury bude možné rozhodnout, jestli alespoň v principu takové schopnosti mohou mít.

Poděkování: Tato práce vznikla v rámci výzkumného záměru AV0Z10300504 s částečnou podporou grantu 1ET100300419

Literatura

- [1] I. Aleksander, B. Dummall: Axioms and Tests for the Presence of Minimal Consciousness in Agents. *Journal of Consciousness Studies*, Volume 10, No. 4-5, 2003.
- [2] M. A. Arbib: The Mirror System Hypothesis: How did protolanguage evolve? In: Maggie Tallerman, editor, *Language Origins: Perspectives on Evolution*. Oxford University Press, 2005.

- [3] M. Blum, R. Williams, B. Juba, M. Humphrey: Toward a High-level Definition of Consciousness, Invited Talk to the *Annual IEEE Computational Complexity Conference*, San Jose CA, (June 2005).
- [4] R.A. Brooks: Intelligence without reason. *Proceedings of the 12th Intl. Conference on Artificial Intelligence (IJCAI-91)*, 1991, pp. 569-595
- [5] H. Cruse: The evolution of cognition— a hypothesis. *Cognitive Science* Vol. 27 No. 1, 2003, pp. 135-155.
- [6] D. Dennett: *Consciousness Explained*, The Penguin Press, 1991.
- [7] J. Feldman: *From Molecule to Metaphor*. MIT Press, Cambridge, MA, June 2006.
- [8] V. Gallese, M. E. Eagle, P. Migone: Intentional attunement: Mirror neurons and the neural underpinnings of interpersonal relations. *J. of the American Psychoanalytic Association*, 55: 131-176, 2007.
- [9] S. Harnad: The Symbol Grounding Problem. *Physica D* 42: 335-346, 1990.
- [10] O. Holland (Editor): *Journal of Consciousness Studies*, Special Issue: *Machine Consciousness*. Volume 10, No. 4-5, April-May 2003.
- [11] O. Holland, R. Goodman: Robots With Internal Models: A Route to Machine Consciousness? *Journal of Consciousness Studies* Volume 10, No. 4-5, April-May 2003.
- [12] D. Hume: Enquiry Concerning Human Understanding, in *Enquiries concerning Human Understanding and concerning the Principles of Morals*, edited by L. A. Selby-Bigge, 3rd edition revised by P. H. Nidditch, Oxford: Clarendon Press, 1975.
- [13] J. R. Hurford: Language beyond our grasp: what mirror neurons can, and cannot, do for language evolution. In: O. Kimbrough, U. Griebel, K. Plunkett (eds.): *The Evolution of Communication systems: A Comparative Approach*. The Vienna Series in Theoretical Biology, MIT Press Cambridge, MA, 2002.
- [14] M. Minsky: Consciousness is a big suitcase. EDGE, http://www.edge.org/3rd_culture/minsky/minsky_p2.html, 1998
- [15] M. Novák: Každý sám svým páněmbohem. *Světová literatura* r. 10, č. 4, s. 165-194, 1965.
- [16] R. Pfeifer, J. Bongard: *How the body shapes the way we think: a new view of intelligence*. The MIT Press, 2006.
- [17] R. Pfeifer, C. Scheier: *Understanding Intelligence*. The MIT Press, Cambridge, Massachusetts, London, England, 1999, 697 s.
- [18] V. S Ramachandran: Mirror neurons and imitation as the driving force behind “the great leap forward” in human evolution. *EDGE: The third culture*, viz http://www.edge.org/3rd_culture/ramachandran/ramachandran_p1.html
- [19] V. S. Ramachandran: Mirror neurons and the brain in the vat. http://www.edge.org/3rd_culture/ramachandran06/ramachandran06_index.html
- [20] G. Rizzolatti, L. Fadiga, V. Gallese, I. Fogassi: Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, 3:131-141,1996.
- [21] A. Smee: *Principles of the Human Mind Deduced from Physical Laws*, 1849 (N.Y. 1853).
- [22] J. R. Searle: *Minds, Brains, and Programs*. Behavioral and Brain Sciences, Vol 3, No. 3, s. 169-225, 1980.
- [23] L. L. Steels, M. Loetzsch, M. S. Spranger: *Semiotic Dynamics Solves the Symbol Grounding Problem*. Available from *Nature Precedings* <http://hdl.nature.com/10101/npre.2007.1234.1>, 1997
- [24] A. Turing: Computing machinery and intelligence. *Mind*, vol. LIX, no. 236, October 1950, pp. 433-460.
- [25] L. G. Valiant: *Circuits of the Mind*. Oxford University Press, New York, Oxford, 1994, 237 p.
- [26] J. Wiedermann.: *Towards Algorithmic Explanation of Mind Evolution and*

- Functioning (Invited Talk). In: L. Brim, J. Gruska and J. Zlatuška (Eds.), *Mathematical Foundations of Computer Science, Proc. of the 23-rd International Symposium (MFCS'98)*, Lecture Notes in Computer Science Vol. 1450, Springer Verlag, Berlin, 1998, pp. 152—166.
- [27] J. Wiedermann: Mirror Neurons, Embodied Cognitive Agents and Imitation Learning. In: *Computing and Informatics*. Vol. 22, no. 6 (2003), p. 545-559.
- [28] J. Wiedermann: Chtěli byste být mozem v baňce? *Pokroky matematiky, fyziky a astronomie*, 51(4), str. 272-282, 2006.
- [29] J. Wiedermann: HUGO: A Cognitive Architecture with an Incorporated World Model. *Proc. of the European Conference on Complex Systems ECCS'06*, Saïd Business School, Oxford University, 2006, viz také Technical report No. 966, Ústav informatiky AV ČR, 2006.

Modely mysle

Copyright © Jan Burian, Igor Farkaš, Jana Horáková, Jozef Kelemen, Juraj Hvorecký, Alica Kelemenová, Ladislav Kováč, Vladimír Kvasnička, Mária Markošová, Jiří Pospíchal, Karel Pstružina, Ján Rybár, Ján Šefránek, Radovan Šikl, Jiří Wiedermann 2008

Published by © Vydavateľstvo Európa, s.r.o., Bratislava 2008.

Illustrations © Vladimír Havrilla 2008

Cover design and layout © Bee and Honey 2008.

First Slovak edition. Pages 248.

All rights reserved. This publication may not be reproduced, stored in retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise without the prior permission of the copyright owner.



Europa Publishing House

Podháj 15

841 03 Bratislava

Slovak Republic

europa@eastfund.sk

www.vydavatelstvo-europa.sk

ISBN 978-80-89111-34-3

Copyright © Jan Burian, Igor Farkaš, Jana Horáková, Jozef Kelemen, Juraj Hvorecký, Alica Kelemenová, Ladislav Kováč, Vladimír Kvasnička, Mária Markošová, Jiří Pospíchal, Karel Pstružina, Ján Rybár, Ján Šefránek, Radovan Šikl, Jiří Wiedermann 2008

Ako piaty zväzok edície Šimon vydalo © Vydavateľstvo Európa, s.r.o., Bratislava 2008.

Vydavateľ Martin Plch. Prvé slovenské vydanie. Počet strán 248.

Návrh obálky, grafická úprava a sadzba © Bee and Honey 2008.

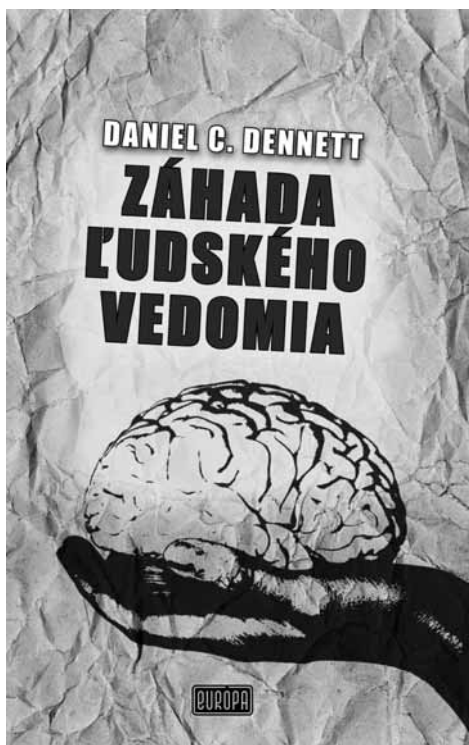
Korektúry Viera Juričková.

Tlač – Tlačiareň P+M Martin.

Všetky práva vyhradené. Nijaká časť tejto knihy nesmie byť reprodukováaná, uchovávaná v rešeršných systémoch, alebo prenášaná akýmkoľvek spôsobom vrátane elektronického, mechanického, fotografického či iného záznamu bez predchádzajúceho písomného súhlasu majiteľov autorských práv.

V edícii Šimon vyšlo

Michaela Martinková – Chlapec, ktorý myslí v obrazoch a vyjadruje sa kresbou, alebo sprevádzanie malého nadaného autistu



Predslov: Egon Gál

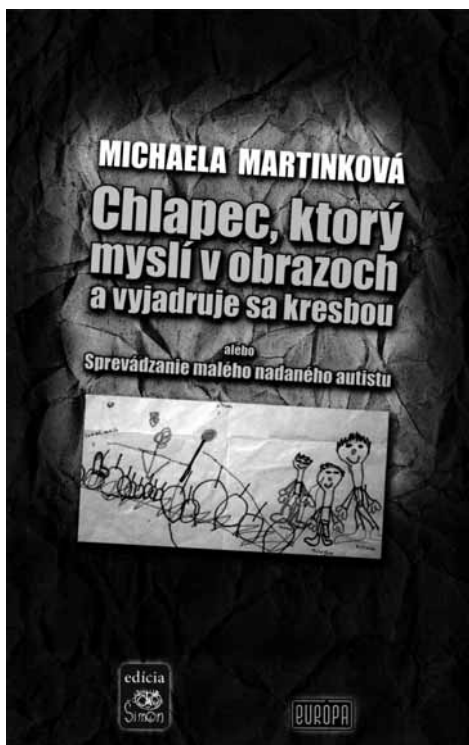
Doslov: Andrea Šedibová

Autistické dieťa sa nevie samé hrať, resp. jeho spontánna aktivita ho postupom času nesocializuje, ale práve naopak – odvádza ďalej od sociálne akceptovateľného správania, ktorým by sa dieťa mohlo zaradiť do života rodiny a spoločnosti. Je úplne neschopné alebo má významne zníženú schopnosť postarať sa samé o seba, samostatne si organizovať život, tvorivo, konštruktívne a sociálne prijateľne tráviť voľný čas, zapájať sa do organizovaných sociálnych aktivít intaktnej populácie, respektíve rovesníkov a vlastnej rodiny. Je nesmierne náročné porozumieť uzavretému svetu autizmu, nájsť most, ktorým by sme objavovali cestu k porozumeniu a spoznaniu svojského sveta každého autistu.

Výnimočná kniha - denník „profesionálnej matky“ nadaného autistického chlapca – je prvým edičným počínom, zoznamujúcim slovenských čitateľov s problémom autistických detí.

Nicholas Humphrey – Vnútorné oko

Preklad: Martin Jankovič



Knih je prepracovanou verziou súboru televíznych prednášok o evolúcii sociálnej inteligencie, ktoré Nicholas Humphrey predniesol roku 1986 v televízii BBC. Jej ústrednou témou je problém vedomia a jeho úlohy v sociálnom živote. V jednotlivých kapitolách sa autor zamýšľa nad otázkami: Čo je vedomie? Ako funguje? Ako ho možno poznávať? Čím sa myseľ človeka odlišuje od mysle iných živočíchov?

Prístupným a zasväteným spôsobom píše Humphrey o najnovších poznatkoch kognitívnej psychológie a filozofie mysle i o výsledkoch vlastných výskumov života goríl v Rwande aj o svojich psychologických experimentoch. V tejto originálnej a nesmierne pútavej publikácii po prvý raz sformuloval hypotézu, podľa ktorej „vnútornému oku“ – schopnosti nazrieť do vlastnej mysle, aj mysle iných ľudí –, vďačíme za sociálnu inteligenciu, schopnosť empatie, morálky, aj za schopnosť byť krutý.

Daniel C. Dennett – Záhada ľudského vedomia
Preklad: Richard Cedzo



Daniel C. Dennett je jeden z najprovokatívnejších a najinšpiratívnejších súčasných filozofov zaoberajúcich sa ľudským myslením a vedomím. Jeho myšlienkami sa inšpirujú ľudia pracujúci v rôznych oblastiach sociálnych a humanitných vied (psychológovia, sociológovia, kultúrni antropológovia, religionisti). Je príkladom „interdisciplinárneho“ filozofa, ktorý má hlboké vedomosti aj z mimofilozofických oblastí, akými sú kognitívna veda, umelá inteligencia, evolučná teória a pod. Je príkladom moderného filozofa, ktorý okrem toho, že je schopný kompetentne sa vyjadrovať k problémom filozofie mysle a ľudského vedomia, je nesmierne inšpiratívny aj pre laikov zaujímajúcich sa o problémy myslenia a vedomia.