Edícia výskumných textov
informatiky a informačných technológií

Myseľ, inteligencia a život

Vydané s podporou firmy Gratex International, a.s.

Vladimír Kvasnička
Peter Trebatický
Jiří Pospíchal
Jozef Kelemen
(editori)

Myseľ, inteligencia a život

Slovenská technická univerzita v Bratislave, 2007

Edícia výskumných textov informatiky a informačných technológií

Myseľ, inteligencia a život

Vydané s podporou firmy Gratex International, a.s.

© FIIT STU, 2007

Cover Designer: Peter Kaminský, s použitím obrázku J. Čapka

Vydala Slovenská technická univerzita v Bratislave
vo Vydavateľstve STU, Bratislava, Vazovova 5.

ISBN

EDÍCIA VÝSKUMNÝCH TEXTOV
INFORMATIKY A INFORMAČNÝCH TECHNOLÓGIÍ

Vladimír Kvasnička, Peter Trebatický, Jiří Pospíchal a Jozef Kelemen, editori

MYSEĽ, INTELIGENCIA A ŽIVOT

1. vydanie
Tlač Vydavateľstvo STU v Bratislave

521 strán

2007

ISBN

Obsah

Predhovor vii

Časť I – Filozofia mysle

Hľadanie kauzálnych vzťahov v probléme mysle a tela z pohľadu neredukcionistického fyzikalizmu 3
Igor Farkaš

Tělo a mysl – východiska a alternativy 17
László Hajnal

Prožívání epizodických situací 27
Ivan M. Havel

Cesta ke kyborgovi 71
Jana Horáková, Jozef Kelemen

Intencionalita zážitkov 89
Juraj Hvorecký

O vzťahu prírodných a spoločenských vied 95
Ladislav Kvasz

Abstrakce, komparace a generalizace jako základní operace lidského myšlení 111
Karel Pstružina

Časť II – Kognitívna veda

Mozgové štruktúry a mechanizmy riadenia strategickej pozornosti 121
Rudolf Andoga, Norbert Kopčo

Kurt Gödel: život, výsledky a jejich význam 129
Libor Běhounek

Neurovedné okno do vedomia 145
Ľubica Beňušková

Význam emocí pro umělé bytosti 157
Michal Bída, Rudolf Kadlec, Cyril Brom

Samoorganizace a kognice 173
Jan Burian

Kognitívne východiská informačnej vedy 191
Mirka Grešková

Racionalita z pohledu logiky 209
Petr Jirků

Pokus o klasifikáciu optických ilúzií 223
Ján Rybár

Kognitívna teória usudzovania 233
Ján Šefránek

Časť III – Umelá inteligencia

Hierarchické zobecňující gramatiky 243
Tomáš Brandejský

Proč je hranice mezi virtuální bytostí a jejím světem neostrá 257
Cyril Brom, Jiří Lukavský

Učení neuronových sítí se schopností generalizace 275
Věra Kůrková

Modelovanie na báze reaktívnych agentov 287
Andrej Lúčny

Jazyk ako graf 297
Mária Markošová, Peter Náther

Kognitivní aspekty znalostního managementu 309
Peter Mikulecký

Zložitosť (kybernetický pohľad) 317
Ján Sarnovský

Myseľ, inteligencia a inteligentné technológie 329
Peter Sinčák

Kognitívna sémantika komplexných kategórií založená na rozlišovacích kritériách . 339
Martin Takáč

Korektúra textu neurónovými sieťami s echo stavmi 357
Peter Trebatický

Časť IV – Umelý život

Život ve světě symbolů: počítání pomocí membrán 369
Luděk Cienciala, Alica Kelemenová

Vplyv mutácie na štruktúru organizmov 389
Marián Ivančo

Lindenmayerove systémy a ich tvorca v spomienkach, faktoch a citátoch 399
Alica Kelemenová

Fundamentálne princípy kognitívnej biológie 415
Ladislav Kováč

Molekulárna darwinovská evolúcia in silico 441
Vladimír Kvasnička

Ontogeneze jako algoritmus a jako výklad 465
Anton Markoš, Lenka Sýkorová, Libor Sýkora

Evolúcia jazyka a univerzálny darwinizmus 475
Jiří Pospíchal, Vladimír Kvasnička

Spojení samoorganizace s výpočty: minimální život v moři umělých molekul 497
Jiří Wiedermann

Index autorov 513

Hľadanie kauzálnych vzťahov v probléme mysle a tela z pohľadu neredukcionistického fyzikalizmu

Igor Farkaš[1]

Abstrakt. Tento článok ponúka náčrt výpočtového modelu vzťahu myseľ-mozog, ktorý sa opiera o poznatky z kognitívnej neurovedy, teórie nelineárnych dynamických systémov a samo-organizácie, pričom sa kladie dôraz na vysvetlenie kauzálnej interakcie medzi mozgom a mysľou. Budem argumentovať, že na lepšie pochopenie tohto problému je nutné uvažovať tri úrovne opisu reality na rozdiel od tradičných dvoch, pričom kauzálne princípy sú realizované v rámci fyzikálneho systému. Mentálna úroveň opisu však ostáva užitočnou, dokonca nevyhnutnou súčasťou modelu, neredukovateľnou na fyzikálnu úroveň. Z pohľadu existujúcej taxonómie filozofických prístupov má prezentovaný model najbližšie k princípom neredukcionistického fyzikalizmu.

1 Úvod

Vzťah mysle a tela je všeobecne považovaný za centrálny problém filozofie mysle, aj keď pozornosť pútajú aj iné otázky zamerané na povahu „mentálnosti“ bez hľadania vzťahu s fyzikálnym svetom [38]. Problém vzťahu mysle a tela bol od antických čias výsostnou doménou filozofov, no v polovici minulého storočia skúmanie tohto zložitého filozofického a vedeckého problému nadobudlo interdisciplinárny charakter, pretože sa prenieslo aj do portfólia psychológov, neurobiológov či informatikov. Principiálna otázka sa týka vzťahu medzi fyzikálnymi udalosťami, o ktorých hovoríme v kontexte mozgu a mentálnymi udalosťami, ktoré spájame s mysľou. Dualizmus a monizmus predstavujú dva základné myšlienkové prúdy, pričom každý má svoje vnútorné členenie [73]. V rámci týchto prúdov však badať nejednoznačnosť, pretože niektoré teórie, najmä tie „neextrémistické,“ ponúkajú viac ako jednu interpretáciu, a ťažko ich zaradiť do jedného či druhého hlavného prúdu.

Dualistické názory, ktorých vznik siaha až k Aristotelovi v západnej kultúre, no taktiež k rôznym východným filozofiám, postulujú existenciu dvoch odlišných svetov alebo úrovní reality. Exaktnejšiu podobu tomuto myšlienkovému prúdu dal Descartes, ktorý tvrdil, že tieto dva svety sa kvalitatívne líšia, pretože materiálne entity sú priestorové, majú vlastnosti a sú objektívne pozorovateľné, zatiaľ čo mentálne entity nie sú priestorové, nemajú vlastnosti, a sú vnútorne prístupné len samotnému subjektu [16]. Ešte môžeme dodať, že fyzikálne procesy sú samy osebe bez významového obsahu a sú nevedomé, zatiaľ čo mentálne procesy sú intencionálne (sú o niečom, zamerané na niečo) a môžu byť vedomé. Descartes bol zástancom viery v paralelnú existenciu dvoch svetov – *dualizmus substancií*, zatiaľ čo iný dualistický smer – *interakcionizmus* – postuluje obojsmerné kauzálne vzťahy medzi oboma svetmi [50]. Dualistické názory majú svoje opodstatnenie a historické dôvody prameniace zo západných a východných svetonázorov. Okrem toho, argumenty pre ne možno hľadať v subjektívnosti zážitkov človeka

[1] Katedra aplikovanej informatiky FMFI, Univerzita Komenského, Mlynská dolina, 842 48 Bratislava. Autor pracuje na čiastočný úväzok i na Ústave merania SAV v Bratislave. E-mail: farkas@fmph.uniba.sk

i v samotnom fakte, že okolitý svet je plný dichotómií. Naše vlastné empirické skúsenosti nás, zdá sa, nabádajú k dualizmu: Už dávno sa vedelo, že poškodenia v mozgu vedú k duševným poruchám, a opačne, poznáme napr. pozitívne účinky psychoterapie v lekárskej praxi. „Extrémistické“ formy dualizmu (ako napr. dualizmus substancií) sú však problematické, pretože kvôli kvalitatívnym rozdielom medzi oboma svetmi nedokážeme pomocou súčasných vedeckých metód vysvetliť ich vzájomné interagovanie. Menej radikálne formy dualizmu postulujú, že ide o jednu fyzikálnu realitu, ale na jej úplné pochopenie je treba dve rôzne úrovne opisu v rôznych perspektívach. *Dualizmus vlastností* predpokladá, že mentálne vlastnosti mozgu nie sú redukovateľné na fyzikálnu úroveň, aj keď z nej emergujú. Trochu odlišne vyznieva *explanačný dualizmus* populárny najmä v psychiatrii [37], ktorý zdôrazňuje, že na úplné pochopenie človeka sú nutné oba spôsoby vysvetlenia, pričom neuvažujeme žiadne predpoklady o ich vzájomnom vzťahu [47].[2]

Ako protipól k dualizmu tradične stoja *monistické* názory, ktoré postulujú existenciu len jedného sveta, jednej reality. Na jednej strane stojí *idealizmus,* ktorého základnou premisou je, že jedinou realitou sú mentálne entity (napr. v duchu tvrdenia „existovať znamená byť vnímaný”). Aj keď tento pohľad má tiež svojich zástancov, významnejšie postavenie má opačne stojaci *fyzikalizmus* (resp. materializmus), a to hlavne hlavne v západnej filozofii. Podľa fyzikalizmu jedinou existujúcou entitou je telo (mozog), pričom vzťah mysle k nemu môže mať rôzne podoby [62]. *Redukcionistický fyzikalizmus*, tiež známy ako eliminačný materializmus, chápe mentálne stavy už ako epifenomén (t. j. sprievodné mentálne javy sú subjektívne vnímané jedincom, sú však redukovateľné na opis mozgu) a predikuje, že v konečnom dôsledku bude mentálna úroveň reality redundantná a úplne vysvetliteľná na fyzikálnej úrovni mozgu. Takýto redukcionistický pohľad má aj v súčasnosti svojich zástancov, no ostáva problematickým, pretože v absolutistickom chápaní smeruje k redukcii ad absurdum [10]. Kompromisnejšiu formu ponúka viac uznávaný *neredukcionistický fyzikalizmus*, ktorý predpokladá, že fyzikálny svet je kauzálne uzavretý, a podľa ktorého mentálne stavy sú síce závislé od fyzikálnej entity, avšak ich opis (epistemologické poznanie) a vysvetlenie nie sú redukovateľné na opis jazykom fyzikálnej vedy. Tento filozofický smer (ktorý má svoje vnútorné členenie) nadobudol na popularite po tom, čo bol zavedený koncept *superveniencie*, podľa ktorého mentálne vlastnosti sú akousi „pridanou hodnotou“ k fyzikálnym vlastnostiam, pričom nie sú na ne redukovateľné, a platí princíp viacnásobnej realizovateľnosti mentálnych vlastností [14].[3]

Viacnásobná realizovateľnosť vyplýva aj z modernejšej monistickej doktríny – *funkcionalizmu,* ktorý sa zrodil v súvislosti so vznikom počítačov. Funkcionalizmus hovorí o mentálnych stavoch (alebo udalostiach) nezávisle od ich implementácie (softvér nezávislý od hardvéru), len v kontexte ich funkcie v rámci fyzikálneho systému. Viacnásobnú realizovateľnosť pritom nezužuje na rôzne konfigurácie toho istého substrátu (rôznych fyzikálnych stavov mozgu), ale zahŕňa aj iné, napr. nebiologické fyzikálne systémy (vrátane počítačov), v ktorých môžu vznikať rovnaké mentálne stavy, ak budú plniť rovnakú funkciu v systéme. Voči funkcionalizmu boli postupne vznesené námietky

[2] Ide teda o dva spôsoby (dve ontológie) pochopenia, opisu jednej reality, na rozdiel od substančného dualizmu, kde sa postulujú dve (ontologicky) odlišné reality.

[3] Vzťah superveniencie medzi fyzikálnymi vlastnosťami (FV) i mentálnymi vlastnosťami (MV) je asymetrický: ak sa u dvoch jedincov líšia FV, líšia sa aj MV, naopak to nemusí platiť.

najmä vo forme myšlienkových experimentov, ktoré nabádajú, že funkcionalizmus je nedostatočným vysvetlením vzťahov medzi mentálnym a fyzikálnym svetom. Napríklad argument známy ako problém invertovaného spektra implikuje, že môžu existovať dvaja ľudia s rovnakými funkcionálnymi stavmi, ale rôznymi mentálnymi stavmi [3]. Podobne, problém čínskej izby zase implikuje, že keďže mentálne stavy musia zahŕňať aj sémantickú zložku, čisto syntaktické manipulácie nestačia, aj keď sú funkcionálne plnohodnotné [57]. Sám Searle [58] prišiel s vlastnou monistickou teóriou nazvanou *biologický naturalizmus*, podľa ktorej všetky mentálne javy (od bolesti po až po abstraktnejšie myšlienky) sú spôsobené neurobiologickými procesmi v mozgu, a že súčasne mentálne stavy a procesy sú vlastnosťami mozgu na vyššej úrovni opisu. Hovorí teda o dvoch ontológiách – prvej a tretej osoby, ktoré sa obe viažu na tú istú fyzikálnu realitu – mozog.[4] Mentálne vlastnosti mozgu niektorí interpretujú ako prejavy *emergencie*, čo je veľmi populárna doktrína, ktorá v teórii komplexných systémov a samoorganizácie (v rôznych doménach) predstavuje ich kľúčovú vlastnosť [28].

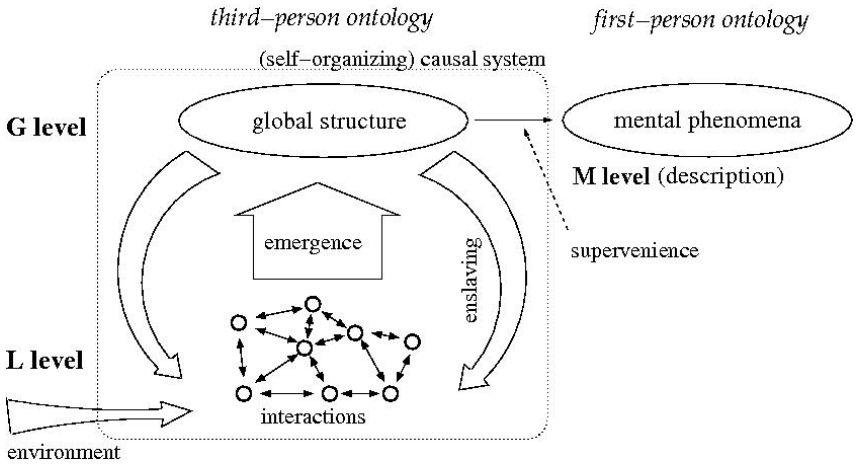
Je zrejmé, že vo filozofii problému vzťahu myseľ-mozog neexistuje konsenzus. Samotný Searle [58] dokonca kritizuje historicky vzniknuté dichotomické vnímanie tohto problému, s čím sympatizujeme. Okrem toho, čo vo väčšine prezentovaných prístupov zjavne absentuje, je to, že ak teda predpokladáme existenciu dvoch zásadne (ontologicky) odlišných druhov entít, či úrovní reality, ako potom tieto kauzálne interagujú, a to oboma smermi. Emergencia ponúka možné vysvetlenie kauzality zdola nahor. Problém potenciálneho kauzálneho účinku mentálnych javov na fyzikálnu substanciu (t. j. účinku zhora nadol) – problém *mentálnej kauzality* sa v literatúre tiež objavuje v rôznych podobách [52]. Viacero filozofov upustilo od uvažovania o nemateriálnej povahe mysle a stotožnilo sa najmä so smerom neredukcionistického fyzikalizmu, ktorý nevylučuje možnosť dvoch ontológií – rôznych jazykov opisu činnosti mozgu, pri zachovaní monistického pohľadu na problém. Tým sa však problém mentálnej kauzality nerieši, a dokonca vzniká protirečenie: ako môžu byť mentálne javy/vlastnosti kauzálne relevantné, keď fyzikálny svet je, podľa predpokladov neredukcionistického fyzikalizmu, kauzálne uzavretý? (O problémoch mentálnej kauzality a rôznych pohľadoch podrobne pojednáva [40], pozri aj [69]).

Naším presvedčením je to, že schodnou cestou k riešeniu problému vzťahu mozog-myseľ je orientácia na výpočtové modely, ktoré sú explicitnejšie než verbálne filozofické teórie, aj keď sa snažia „len“ formálne opísať nejaký vzťah. Dávajú však podľa nás priestor pre pochopenie obojsmernej kauzality. V ďalšej časti predstavíme náš model, spomenieme podporné argumenty preň a pokúsime sa ho podľa jeho vlastností zaradiť do existujúcej taxonómie vyššie spomínaných filozofických teórií. V poslednej časti pojednávame o implikáciách nášho prístupu.

[4] Podobne ako iné názory, aj tento má svojich oponentov (napr. [38]), ktorí kritizujú vnútornú nekonzistentnosť Searlovej teórie.

2 Model vzťahu mozog-myseľ

Pri opise nášho modelu najprv spomenieme jeho základné vlastnosti a koncepty, na ktorých bude postavený. Náš model je: (1) výpočtový, (2) založený na samoorganizácii [21], a snaží sa opísať (3) obojsmernú kauzalitu medzi „mentálnym“ a fyzickým, opierajúc sa o poznatky z kognitívnej neurovedy, systémov so samoorganizáciou a z teórie nelineárnych dynamických systémov [20]. V ďalšom texte tieto vlastnosti, koncepty a vzájomné vzťahy bližšie opíšeme.



Obrázok 1. Ponúknutý formálny model vzťahu mozog-myseľ pozostávajúci z troch úrovní opisu. Vzťahy susedných úrovní 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 z perspektívy prvej osoby. Fyzikálny systém je otvorený, prijíma externé stimuly na úrovni L.

Schéma najjednoduchšieho možného modelu je znázornená na obr. 1. Model pozostáva z *troch* (namiesto tradičných dvoch) úrovní opisu reality (dvoch fyzikálnych a jednej mentálnej), ktoré sú nevyhnutné na pochopenie problému myseľ-mozog, a ktoré sú konzistentné s troma hore uvedenými bodmi.[5] Výpočtový model je založený na predpoklade, že správanie ľubovoľného existujúceho systému (alebo opatrnejšie povedané, aspoň niektoré jeho aspekty) je možné formálne opísať. Je zrejmé, že týmto sa naše možnosti viac-menej zužujú na perspektívu tretej osoby, a teda že uvažované úrovne opisu v modeli sa týkajú mozgu (mentálnej úrovni opisu však tiež budeme venovať pozornosť). Pre naše potreby postačí, ak za najelementárnejšiu fyzikálnu úroveň opisu budeme považovať biologické neróny 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 Broadmannove kortikálne arey), i s ich vzájomnými interakciami. Táto vlastnosť hierarchického prepojenia a ovplyvňovania susedných úrovní je pre mozog signifikantná na rozdiel od počítača [5], no budeme od tohto faktu abstrahovať, a teda predpokladať

[5] 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 problému myseľ-mozog z pohľadu prezentovaného modelu. Pre zjednodušenie slovníka budeme aj M chápať ako úroveň opisu (aj keď tá je škálovo identická s G).

existenciu len dvoch interagujúcich fyzikálnych úrovní, lokálnej L (častí mozgu) a globálnej G (mozog ako celok). Pritom treba zdôrazniť, že obe fyzikálne úrovne sú relatívne v tom zmysle, že fyzikálna škála, ktorej sa budú týkať, bude závisieť od kontextu. Napríklad, úroveň neurálnych zhlukov je globálna v kontexte úrovne neurónov, ale lokálna vo vzťahu ku kortikálnym areám. Túto abstrakciu si berieme na pomoc preto, lebo existujúce empirické a teoretické argumenty, ktoré spomenieme, existujú na rôznych škálach, no v podstate sa dajú chápať ako ten istý princíp kauzality.

V modeli figuruje i mentálna úroveň opisu M, ktorú dáme do vzťahu s fyzikálnou úrovňou G. Úroveň M priamo v modeli nevystupuje ako súčasť kauzálneho systému, o ktorom predpokladáme, že je kauzálne uzavretý,[6] ale len „prostredníctvom“ úrovne G. Preto budeme hovoriť o „mentálnej“ kauzalite majúc na pamäti, že „strojcom“ kauzality zhora-nadol je samotný mozog, pričom mentálne stavy v mysli sú opisom globálnych stavov mozgu z pohľadu prvej osoby. Nie je však potom opis vlastností mozgu na úrovni M redundantný, keď sa tieto nepodieľajú na riadení jeho činnosti? Odpoveď ponúkneme v závere článku, potom, čo detailnejšie nazrieme do vlastností modelu.

3 Hľadanie oporných argumentov

Oporné argumenty ponúkneme z oblasti kognitívnej a výpočtovej neurovedy, samoorganizácie, a teórie nelineárnych dynamických systémov.

3.1 Čo hovorí kognitívna neuroveda

V oblasti kognitívnej neurovedy sa pozornosť sústreďuje na hľadanie tzv. *neurálnych korelátov* mentálnych stavov, ktoré sa dajú neinvazívne merať vďaka existujúcim technikám zobrazovania činnosti mozgu v priestore i čase [23]. V súčasnosti existuje pomerne dosť empirických výsledkov získaných zo zvieracích alebo ľudských subjektov, ktoré poukazujú na to, že kognitívne procesy sú sprevádzané prechodnou integráciou viacerých distribuovaných interagujúcich oblastí v mozgu (napr. [13,65,67,19,6]. Najprijateľnejším mechanizmom umožňujúcim takúto integráciu je vytvorenie dynamických spojení pomocou prechodne *synchronizovanej aktivity* neurónov v rôznych frekvenčných pásmach. Synchronizácia tu znamená pálenie neurónov s rovnakou fázou alebo v presnom fázovom závese, čo sa dá merať a kvantifikovať modernými štatistickými metódami [42]. Synchronizovaná aktivita bola pozorovaná v rôznych priestorových škálach. Krátkodosahové synchronizácie sa týkajú jednej senzorickej modality a sú zrejme potrebné na perceptuálne viazanie príznakov (feature binding) vnímaného objektu, aby mohol vzniknúť koherentný vnem [29].

Boli však pozorované i ďalekodosahové synchronizácie, ktoré zahŕňajú vzdialené kortikálne oblasti, a ktoré sprevádzajú rôzne kognitívne úlohy, napr. rozpoznanie vizuálneho vnemu spojené s odpovedajúcou motorickou reakciou. V pozoruhodnom experimente s Mooneyho tvárami[7] [53] boli asi štvrtinu sekundy po ukázaní podnetu pozorované kolektívne gama oscilácie neurónov (40 Hz) vo viacerých oblastiach (čo odpovedalo zrejme spracovaniu vizuálneho podnetu), avšak synchronizácia nastala iba v prípade, ak sa subjekt díval na tvár. Druhý vzorec synchronizácie bol potom pozorovaný

[6] V zmysle slabej definície: „Žiaden fyzikálny jav nemá príčinu mimo fyzikálnej domény.“ (J. Kim)
[7] Profily abstraktných čiernobielych tvárí v kontraste s tými istými podnetmi obrátenými dolu hlavou.

v motorickej kôre, čo korelovalo s motorickou odpoveďou (kliknutie myšou) a ten pochopiteľne v oboch prípadoch. V inom zaujímavom experimente bola daná do súvisu synchronizácia s kognitívnym procesom – krátkodobou pamäťou [22]: iba ak nastala synchronizácia (v hipokampe a priľahlých štruktúrach) v čase memorovania, po ktorom nasledovalo zámerné rozptýlenie, bol subjekt schopný neskôr si vyvolať z pamäti zapamätané podnety. Tieto výsledky sú kompatibilné aj so všeobecnejšou predstavou o synchronizácii, ktorá zahŕňa všetky dimenzie kognitívneho aktu, vrátane asociatívnej pamäti, emocionálneho podfarbenia a motorického plánovania [13,68]. Zdôraznenie synchrónnej aktivity implikuje, že odpoveď na podstatu korelátov mentálnych stavov treba hľadať nie v štrukturálnych vzťahoch aktivovaných oblastí mozgu, ale v ich *dynamickom vzorci správania*. Tieto vzorce vznikajú v časovom rozpätí rádovo stoviek milisekúnd, čo je čas potrebný na vznik kognitívneho vnemu. Inými slovami, tieto poznatky z neinvazívnych metód napovedajú, že je trochu problematické zosúvzťažniť fyzikálne a mentálne stavy (ako sa bežne uvažuje v dualistických teóriách), pretože obe úrovne operujú na rôznych časových škálach. Napriek tomu, odpoveďou by mohla byť superveniencia mentálnych vlastností v tom zmysle, že mentálny stav odpovedá fyzikálnemu procesu (sekvencii fyzikálnych stavov), pričom platí princíp viacnásobnej realizovateľnosti.

Zaujímavé poznatky nám poskytujú aj invazívne metódy merania aktivity jednotlivých neurónov. Z mozgu sa už napríklad dá vyčítať, ktorý objekt dotyčný subjekt vníma. U makakov bola počas 12.5ms meraná aktivita asi 100 náhodne vybratých neurónov v inferiórnej časti spánkového laloka, a tá robustne a spoľahlivo (pre rôzne objekty, ich rôzne pozície i veľkosti) umožňovala určiť identitu objektu [32].

Hľadanie neurálnych korelátov je skromnejší zámer, ktorý nič nehovorí o ich prípadnom účinku týchto neurálnych korelátov na mentálnu úroveň.[8] Ak by sme chceli vysvetliť tento účinok, museli by sme identifikovať neurálne stavy či procesy, ktoré sú nevyhnutné pre vznik mentálnych stavov, a ktoré ich kauzálne ovplyvňujú. Na to by sme však potrebovali uskutočniť experimenty, v ktorých by sme ovplyvňovali synchronizáciu neurónov a sledovali zmeny v správaní jedinca. Experimenty tohto druhu sa však už dajú vypátrať v literatúre. V prípade hmyzu bolo ukázané, že narušenie synchronizácie spôsobí jeho zníženú schopnosť rozoznávať vône [63]. Zaujímavejší, nedávny experiment s makakmi pomocou zaužívaných behaviorálnych metód ukázal, že stimulácia správnych neurónov reagujúcich na tváre (v inferiórnej časti spánkového laloka) spôsobí, že makak začne vnímať tvár, aj keď samozrejme pred ním žiadna tvár nie je [17].

V kontexte nášho modelu sme sa v ostatných dvoch príkladoch zmienili o experimentoch, ktoré by mohli predstavovať „kauzálny" účinok úrovne G na M. Do úvodzoviek sme ho dali preto, lebo v striktnom slova zmysle ide o metaforu, nakoľko my chápeme možné účinky kauzality len vo fyzikálnom svete, ktorý podlieha fyzikálnym zákonom. Môžeme to teda interpretovať tak, že ide o jav, ktorý nie je formálnym modelom podchytiteľný. Tento jav znamená, že existujúce vlastnosti na úrovni G (napr. synchronizácia) umožnia vznik vlastností mozgu na úrovni opisu M (uvedomenie si pozorovaného objektu), čo možno podľa nás chápať ako supervenienciu. Podobne by bolo

[8] Zámerne sa tu vyhýbame slovu „kauzálnych", aby nevznikla nekonzistentnosť s naším modelom, kde kauzálnosť figuruje len medzi L a G. Tieto účinky na M môžeme chápať ako vznik vlastností M po tom, čo vznikli vlastnosti na úrovni G.

obtiažne hovoriť o mentálnej kauzalite (M→G), pretože tá sa vymyká nášmu modelu. Napriek tomu spomeňme niektoré práce, ktoré nabádajú ku kauzalite zhora nadol.

Existujú experimentálne práce, ktoré poukazujú na to, že stav pozornosti (úroveň M) môže byť akcelerátorom emergentných procesov zdola (L→G). Ako bolo pozorované v experimentoch, stav pozornosti urýchľuje vznik synchronizácie v somatosenzorickej kôre [61] alebo pri vizuálne evokovanej aktivite [24]. Podobne, stav očakávania podnetu môže prispieť k synchronizácii v motorickej kôre [51]. Tieto poznatky vieme vysvetliť pomocou nášho modelu tak, že za štartovací moment (stav pozornosti M) zoberieme jemu odpovedajúci korelát na úrovni G, ktorý potom vyústi v (prostredníctvom procesu L→G) v iný stav G (synchronizácia v kôre).

Ďalší zaujímavý poznatok ponúka predikciu o mentálnej kontrole epilepsie. Je známe, že epileptická aktivita spôsobuje zmeny v mentálnych stavoch, no možno kauzalita funguje i opačne. Už Penfield a Jasper [48] opísali prípad pacienta, u ktorého záchvat v temennom laloku prestal, ak pacient začal riešiť zložitú matematickú úlohu (G→L). Na podobné prípady vplyvu kognitívnych stavov na aktivitu mozgu poukazuje aj Schmidt-Schobstein [56], ktorý predpokladá, že takáto intervencia zhora nadol je možná, nakoľko epileptické zóny sú súčasťou rozsiahlej komplexnej siete pokrývajúcej aj iné oblasti mozgu podieľajúce sa na vytváraní mentálnych stavov. Tým, že jednotlivé oblasti sú vo vzájomnej interakcii, zrejme takýto globálny systém môže mať vplyv na oblasť postihnutú epilepsiou. Tento potenciálny efekt detailnejšie sledovali Lachaux a spol. [42] a ukázali, že percepčné stavy (vznikajúce v mysli subjektu pri riešení úloh) môžu fungovať ako modulátory aktivity neurónov v epileptickom ohnisku. Konkrétne, autori sledovali časové charakteristiky týchto neurónov (dĺžky intervalov medzi výbojmi) a zistili, že tieto sa vplyvom aktu percepcie menili. Konkrétne, aj keď sa na prvý pohľad epileptické ohnisko prejavovalo pravidelnými osciláciami (s frekvenciou 2 Hz) skrytými v gaussovskom šume, následná analýza dát odhalila deterministický vzorec aktivít. Ten sa vplyvom percepčných stavov menil smerom k tzv. nestabilným periodickým orbitám v rámci dynamického systému, čo autori interpretovali ako prejav kauzality zhora nadol. Tento poznatok vznikol kombináciou experimentálnych dát spojených s ich formálnou analýzou, čo nás vedie k druhej doméne výskumu, z ktorej môžeme hľadať podporu pre našu teóriu, a tou sú výpočtové modely so samoorganizáciou, uvedené v ďalšej časti.

Ešte predtým spomeňme príklad kauzality (G→L), ktorá je konzistentná s naším modelom, a to známy fakt týkajúci sa úrovne biologického neurónu. Kortikálny neurón má rádovo 10.000 synaptických spojení od iných neurónov, z čoho vyplýva, že súhra aktivít presynaptických neurónov určuje, ako bude postsynaptický neurón reagovať. Inými slovami, podľa súčasnej predstavy biologický neurón funguje ako integrátor vstupných signálov, ale súčasne aj ako detektor koincidencií na vstupoch (výraznejšie reaguje na súčasný príchod viacerých impulzov v rámci krátkeho časového okna), z čoho vyplýva jeho citlivosť na časové charakteristiky [41].

3.2 Čo hovoria systémy so samoorganizáciou

Samoorganizácia ako proces determinujúci vznik koherentného správania je fascinujúcim fenoménom, ktorý pôsobí ako protiklad k bežne pozorovaným procesom sprevádzaných nárastom neporiadku (entropie) v systéme. Jav samoorganizácie je spoločným znakom rôznych typov prírodných systémov (či už fyzikálnych, chemických alebo biologických),

no príklady takýchto systémov so samoorganizáciou nájdeme aj v sociálnych vedách [31]. Princípy samoorganizácie umožňujú vysvetliť, ako vďaka nelineárnym interakciám medzi elementami (na rôznych škálach) môžu emergovať koherentné vzorce správania až na úrovni celého systému (či už ide o rezonanciu v nejakom fyzikálnom systéme alebo o zosúladený potlesk obecenstva). Formálna analýza takýchto systémov umožnila zodpovedať viaceré otázky a to i vďaka počítačovým simuláciám, ktoré sú kvôli nelineárnym vlastnostiam interakcií často jedinou schodnou cestou, lebo teoretická analýza nie je možná. Aj na mozog sa možno pozerať ako na samoorganizujúci sa systém [70]. Vzniká však otázka, či tento pohľad dokáže uspokojivo vysvetliť po stáročia pertraktovaný vzťah mysle a mozgu. Podľa nášho názoru umožňuje vysvetliť kauzálne vzťahy medzi úrovňami L a G a to hlavne v smere zdola nahor L→G (emergencia). Avšak už Haken [31] poukázal na to, že „zotročovanie“ (enslaving) častí celkom je manifestáciou kauzality zhora nadol, aj keď jej pochopenie je zložitejšie. Tento smer možno ilustrovať na úrovni ľudskej spoločnosti napr. skúsenosťou, že človek má väčšinou tendenciu správať sa podľa davu, či už na ulici alebo v obecenstve podľa potlesku (o zaujímavých príkladoch všadeprítomnej, s tým súvisiacej globálnej koherencie sa možno dočítať v [64]). Dôležitou podmienkou v našom kontexte je to, že prijateľná teória mozgu ako komplexného systému musí uvažovať (predovšetkým) neautonómny režim, t. j. spracovanie vonkajších vstupov (čo v teórii formálnych systémov nie je nutnosťou).[9]

Samoorganizáciu môžeme interpretovať v dvoch rovinách: (1) ako procesy determinujúce kauzálne účinky na úrovni aktivity systému, a (2) ako procesy adaptácie parametrov (plasticity), či už lokálnych (sily prepojení, synáps) alebo globálnych (bude vysvetlené nižšie). Vo výpočtových modeloch môžeme pozorovať účinky samoorganizácie v oboch smeroch, pričom môžeme na rozdiel od vyššie uvedených poznatkov z kognitívnej neurovedy, priamo hovoriť o kauzálnosti, lebo obe úrovne sú kvalitatívne zhodné (fyzikálne), len predstavujú odlišné škály.

Prejavom samoorganizácie pri modelovaní kauzálneho účinku zdola nahor (L→G) pomocou impulzných neurónových sietí (t.j. na úrovni biologických neurónov) je vznik synchronizovanej aktivity v modeli s plasticitou synáps. V súčasnosti najviac experimentálne i teoreticky študovanou formou je plasticita založená na časovaní pre- a postsynaptických impulzov (spike-timing dependent plasticity, STDP; pozri napr. prehľad v [1]), ktorá v modeloch predstavuje hebbovsky založenú adaptáciu [59,45]. Synchronizovanú aktivitu v neurónovej sieti možno pozorovať pri externých štruktúrovaných vstupoch do siete, ale aj pri spontánnych šumových vstupoch [33,26]). Aj na abstraktnejšej úrovni – umelých neurónových sietí bolo navrhnutých zopár hebbovsky založených algoritmov adaptácie váh (synáps), ktoré sú považované za prijateľné modely samoorganizácie v mozgovej kôre (pozri referencie vo [49]). Aj keď boli tieto modely odvodené odlišnými spôsobmi, majú spoločnú tú črtu, že zmenu synaptickej plasticity vzťahujú na postsynaptickú aktivitu nemonotónnym spôsobom. Toto umožňuje dynamickú moduláciu neurónu v procese posilňovania, resp. oslabovania jeho synaptických spojení. Obe kategórie modelov teda ponúkajú mechanizmy samoorganizácie na rôznych úrovniach opisu neurónu z pohľadu diskrétneho času. Nie sú nezlučiteľné, možno medzi nimi nájsť vzťah.

[9] Vychádzame z toho, že bdelý mozog dokáže fungovať i v stave úplnej senzorickej deprivácie, ale len pomerne krátky čas.

Napríklad, BCM pravidlo vyplýva z STDP, ak uvažujeme iba účinky časovo najbližších pre- a postsynaptických impulzov [33].

Čo sa týka kauzality zhora nadol (G→L), zaujímavé predikcie na úrovni neurónu ponúkajú napr. Salinas a Sejnowski [55] v súvislosti s poznatkom o vplyve presynaptických korelácií na správanie neurónu. Výsledky ich modelovania ukazujú, že kľúčom pre zvýšenie temporálneho rozlíšenia v impulznom neuróne je jeho „vyváženosť“, t. j. správny pomer vstupnej excitácie a inhibície. Vyvážený neurón je schopný reagovať aj na impulzné podnety, ktoré sa prejavujú na časovej škále oveľa kratšej než je membránová časová konštanta neurónu (ktorá určuje rýchlosť zmeny membránového napätia neurónu). Inými slovami, vyvážený neurón je oveľa citlivejší na korelácie na vstupoch než nevyvážený neurón, pretože tie ovplyvňujú fluktuácie sumárneho synaptického vstupu, ktorý vybudí neurón k odozve. Autori uvádzajú, že modelovanie vyváženého stavu neurónu bolo vecou nájdenia vhodných parametrov. Zaujímavé by preto bolo zistiť, ako by sa také parametre dali nájsť adaptívne, pomocou mechanizmov samoorganizácie.

3.3 Dynamika kortikálnej koordinácie

Zaujímavé vysvetlenie kauzality zhora nadol na úrovni globálnych aktivít (G) vo formálnom systéme ponúka Kelso [36] pomocou formalizmu nelineárnych dynamických systémov. Na základe experimentálne verifikovaného teoretického modelu navrhol *teóriu dynamiky kortikálnej koordinácie* ako integrujúceho rámca kognitívnych funkcií. Táto teória postuluje existenciu dvoch kľúčových typov premenných.: Jedným sú tzv. *kolektívne premenné,* ktoré globálne špecifikujú vzťahy medzi interagujúcimi zložkami dynamického systému, a usmerňujú ich správanie. Druhým typom sú *riadiace parametre*, ktoré indukujú (fázové) prechody medzi stabilnými stavmi sprevádzané kvalitatívnymi zmenami v správaní systému, a sú tak hybnou silou pre dynamiku vzniku usporiadaných stavov. Smer pôsobenia celku na jeho časti má kvalitatívne odlišný charakter než emergentné prejavy zdola nahor, ktoré spätne pôsobia na celok, a tým aj na tieto premenné (recipročná kauzalita). Identifikácia kolektívnych premenných a riadiacich parametrov v neurálnom systéme je zatiaľ ťažkým orieškom, no v prípade jednoduchých systémov Kelso ponúka vysvetlenie. Napríklad, v experimente s obojručnou koordináciou (pohybujúcich sa ukazovákov ruky subjektu) úlohu kolektívnej premennej spĺňa relatívna fáza medzi oboma prstami, a frekvencia pohybu prstov je riadiacim parametrom [35]. Bolo ukázané, že pre nejaký rozsah frekvencií existujú dve stabilné riešenia (kmitanie prstov vo fáze a v protifáze). Pri náraste frekvencie sa stáva protifáza nestabilnou a nastáva prechod k sfázovanej konfigurácii (zaujímavé je, že opačný prechod nenastáva pri spätnom poklese frekvencie). Napriek tomu, že uvedený experiment je z pohľadu myseľ-mozog triviálny, Bressler a Kelso [7] argumentujú, že kolektívne premenné (napr. relatívna fáza) môžu figurovať aj pri mentálnych procesoch. Podľa teórie dynamiky koordinácie mozog pracuje v metastabilnom dynamickom režime, prechádzajúc permanentne z jedného stabilného režimu do druhého.

Vyššie spomínaný súhrn poznatkov o samoorganizácii predstavuje tehličky do našej neurovýpočtovej mozaiky o tom, ako náš mozog a myseľ asi fungujú. Spomínaná dynamika koordinácie je pozoruhodným kandidátom na vysvetlenie fungovania kortikálnej aktivity na systémovej úrovni. Na neurálnej úrovni zase pomocou modelov s impulznými neurónmi vieme vysvetliť vznik koherentných vzorcov aktivít. Tieto modely síce zahŕňajú aj aspekt plasticity lokálnych parametrov (synaptických spojení,

podľa teórie STDP), no dôležitým krokom by bolo, keby dokázali vysvetliť aj samoorganizáciu globálnych parametrov (ktoré však najprv treba identifikovať). Podľa neurovýpočtového pohľadu je cieľom mozgu ako komplexného systému zrejme jeho maximálna efektívnosť pri reprezentovaní a spracovaní informácie, čo považujeme za konzistentné s požiadavkou na vyváženosť medzi špecializáciou a integráciou jednotlivých centier pri spracovaní informácie v mozgu [66]. Veľmi zaujímavé poznatky ponúkajú nové práce o mozgovej kôre ako o komplexnom systéme s vlastnosťami „malého sveta“, ktorý umožňuje veľmi efektívny prenos informácie v rámci systému (pozri prehľad v [60]). Druhou nedávno objavenou charakteristikou mozgu ako komplexného systému je tzv. bezškálová organizácia [18]. Pohľadu na mozog ako na komplexný systém so samoorganizovanou dynamikou sa venuje i Burian [74].

4 Miesto mentálnej úrovne v modeli

Spomínaný model kladie veľký dôraz na úrovne L a G, a to preto, že len tie sú hybnou silou recipročnej kauzality v mozgu. Znamená to, že za kauzálne účinky považujeme len účinky v rámci prírodných zákonov, jej ostatné prejavy (napr. v makroekonomike) chápeme ako užitočnú metaforu. Za dôležitý však považujeme fakt, že náš formalizmus dáva priestor aj pre „mentálnu“ kauzalitu (t. j. zhora nadol), aj keď na formálnej úrovni sme zatiaľ nedali uspokojivú odpoveď ako k nej dochádza. Podotýkame, že aj mimo rámca formálnych modelov predstavuje mentálna kauzalita zložitý filozofický a vedecký problém [52]. Odhliadnuc od toho vzniká otázka, akú úlohu hrá úroveň opisu M v našom modeli, ak nemá žiadnu kauzálnu silu. Význam úrovne M vidíme v dvoch rovinách.

Po prvé, M predstavuje nevyhnutnú (a tým aj neredukovateľnú) úroveň opisu pre kompletnejšie pochopenie činnosti mozgu. Na to, aby subjekt pochopil mentálne stavy iného subjektu, je nutné skombinovať ontológiu tretej osoby (úroveň G) s ontológiou prvej osoby (M), pričom tu možno uplatniť niečo ako princíp analógie.[10] Týmto je naša pozícia konzistentná s pozíciou neredukcionistického fyzikalizmu. Človeku, ktorý je od narodenia slepý, nemožno vysvetliť, čo je červená farba. Riešením (v budúcnosti) však môže byť vyvolanie tohto zážitku správnou stimuláciou jeho mozgovej kôry (podobne ako v experimente s tvárami, a za predpokladu, že subjekt má funkčnú vizuálnu kôru).

Po druhé, mentálna úroveň opisu mentálnych stavov je prístupná bežnému jazyku a jediná umožňuje efektívnu verbálnu komunikáciu medzi subjektami. Subjekt vlastne ani nemá na výber, pretože úroveň G mu je neprístupná. Súčasne tým umožňuje subjektu figurovať v kontexte populácie, ktorú môžeme chápať ako vyššiu úroveň organizácie s možnými komplexnými prejavmi kolektívneho správania. Aj keby sme v budúcnosti našli presné neurálne koreláty všetkých mentálnych stavov (t. j. že by sme dokázali čítať myseľ), ich opis by bol kvôli ich zložitosti asi prakticky nepoužiteľný.

Náš pohľad môžeme dať aj do súvisu so známymi výpočtovými teóriami *vedomia*. V našom príspevku sme síce zámerne nešpecifikovali, či pod mentálnymi stavmi máme namysli len vedomé alebo aj nevedomé stavy, ale toto rozlíšenie v kontexte recipročnej kauzality nepovažujeme za podstatné. Atkinson a spol. [2] prezentovali taxonómiu

[10] Ja ako subjekt viem porozumieť činnosti mozgu iného subjektu práve preto, že sme z „rovnakého cesta“, aj keď zrejme môžu existovať „reziduálne“ rozdiely v detailoch fenomenálnosti mentálnych stavov (kvôli jedinečnosti osobnej skúsenosti, napr. s vnímaním červenej farby). Týmto súčasne dostávame negatívnu odpoveď na Nagelovu otázku: „Aké je to byť netopierom?“ Nevieme, a ani to nemáme ako zistiť.

výpočtových teórií vedomia podľa dvoch dimenzií.[11] 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 sú jeho koreláty široko distribuované. V rámci tejto taxonómie sa ohľadne prvej dimenzie skôr prikláňame k názoru, že vedomie je špecifickým procesom, avšak ten sa nedá oddeliť od reprezentácií, na ktorých proces operuje (na rozdiel napr. od počítača). To považujeme za konzistentné s požiadavkou ukotvenia neurálnych reprezentácií [72]. Z pohľadu druhej dimenzie je naše zaradenie jednoznačnejšie a to smerom k distribuovanosti neurálnych korelátov vedomia.

5 Implikácie a záver

V prezentovanom modeli chápeme problém vzťahu myseľ-mozog ako problém týkajúci sa komplexného (deterministického) systému so samoorganizáciou, kde neurálne udalosti ovplyvňujú mentálne udalosti a naopak. Touto reciprocitou sa tento prístup odlišuje od „klasických“ fyzikalistických prístupov, ktoré zahŕňajú len neurobiologický prístup zdola nahor (napr. [12]). Obojsmerná kauzalita dáva priestor aj na zásah mentálnych stavov do diania v našom mozgu, ako sme ukázali v predchádzajúcej časti, aj keď kauzálne procesy sa dejú vnútri fyzikálneho sveta.

Jednou z implikácií takéhoto deterministického pohľadu je to, že slobodná vôľa asi predstavuje ilúziu, pretože každý úkon (či už vplyvom zdola alebo zhora) musí byť kauzálne podmienený, a to v rámci fyzikálneho systému. Ilúziu slobodnej vôle chápeme v tom zmysle, že samozrejme *subjekt* je strojcom svojho správania, avšak v rámci neho je to mozog, ktorý riadi činnosť, a vedomie je len „informované“ o dianí.[12] Aj keď takáto hypotéza nie je zrejme veľmi lákavá, v jej prospech existuje viacero experimentálnych poznatkov. Libet [46] síce vo svojich experimentoch, kde ukazuje, že aktivita mozgu predchádza vedomým stavom pri vôľou riadených úkonoch, necháva priestor aj pre zásah slobodnej vôle (a to v podobe vetovacieho mechanizmu, ktorým myseľ zablokuje aktivitu iniciovanú mozgom), avšak viaceré experimenty poukazujú na to, že pocit slobodnej vôle môže byť *iluzórny* (pozri prehľad vo [71]) v tom zmysle, že len úroveň G a externé vstupy určujú správanie subjektu.[13] Táto iluzórnosť môže byť aj dôsledkom nemožnosti našej snahy úplne pochopiť samých seba. V teórii formálnych systémov Gödel [27] dokázal tvrdenie, že žiaden formálny systém nemôže byť kompletný, t. j. že by sa v ňom dali dokázať všetky pravdivé výroky. Ak je mozog možné opísať ako formálny systém, na ktorý Gödelove závery vzťahujú, tak podľa tejto logiky mozog nedokáže sám seba úplne obsiahnuť (pochopiť), a preto niektorým jeho prejavom nemôžeme principiálne porozumieť alebo budeme mať o nich nejaké ilúzie [10].

Táto hypotéza sa môže týkať aj snahy pochopiť zložitejšie aspekty vzťahu mysle a mozgu, ako je napr. fenomenologická povaha mentálnych stavov, vedomia, čo Chalmers [8] definoval ako „ťažké” problémy. Tie vyvolávajú otázku, *prečo* činnosť fyzikálneho systému môže byť spojená so subjektívnym zážitkom, čo implikuje tzv. explanačnú

[11] V kontexte spomínaných neurovýpočtových teórií nie je prekvapením návrh, že by sme mali hľadať nielen neurálne koreláty vedomia, ale aj výpočtové koreláty vedomia, a hľadať medzi nimi vzťah [11].

[12] Máme pri tom na pamäti, že len zlomok procesov prebiehajúcich v mozgu sa dostáva na úroveň vedomia.

[13] Ak pre daný stav v G existuje odpovedajúci stav na úrovni M, vtedy zrejme máme pocit, že sme sa vedome rozhodli.

medzeru medzi funkciou a zážitkom. Mechanistické prístupy (nevynímajúc ten nami prezentovaný) môžu pracovať jedine s postulátom, že fenomenálne javy sú podtriedou vlastností (alebo javmi superveniencie) fyzikálnych systémov (na úrovni G) od istej úrovne komplexnosti. Koho takáto odpoveď neuspokojuje, musí počkať na objavenie nových paradigiem a metód skúmania mysle dúfajúc, že tie mu poskytnú uspokojivejšie odpovede.

Poďakovanie: Tento príspevok vznikol s podporou grantovej agentúry VEGA v rámci projektov 1/2045/05 a 1/3612/06. Autor ďakuje Deziderovi Kamhalovi za cenné pripomienky.

Literatúra

[1] Abbott, L. F., Nelson, S. B.: Synaptic plasticity: Taming the beast. *Nature* 3 (2000) 1178-1183.
[2] Atkinson, A., Thomas, S., Cleeremans, A.: Consciousness: Mapping the theoretical landscape. *Trends in Cognitive Sciences* **4** (2000) 372-382.
[3] Block, N.: Troubles with functionalism. V knihe *Readings in the Philosophy of Psychology*, Block, N. (ed.), Cambridge, MA: Harvard University Press 1980.
[4] Haugeland, J.: Artificial Intelligence. *Behavioral and Brain Sciences* **3** (1980) 219-224.
[5] Bell, A.: Levels and loops: The future of artificial intelligence and neuroscience. *Philosophical Transactions of the Royal Society London B* **354** (1999) 2013-2020.
[6] Beňušková, Ľ.: Kognitívna neuroveda. V knihe *Kognitívne vedy* (Rybár, J., Beňušková, Ľ., Kvasnička, V., eds.). Bratislava: Kalligram 2002, 47-104.
[7] Bressler, S., Kelso, J.: Cortical coordination dynamics and cognition. *Trends in Cognitive Sciences* **5** (2001) 26-36.
[8] Chalmers, D. J.: Facing up to the problem of consciousness. *Journal of Consciousness Studies* **2** (1995) 200-19.
[9] Chrisley, R., Ziemke, T.: Embodiment. *Encyclopedia of Cognitive Science.* Macmillan Publishers 2000, pp. 1102-1108.
[10] Churchland, P. M., Churchland, P. S.: *On the Contrary.* Cambridge: MIT Press 1999.
[11] Cleeremans, A.: Computational correlates of consciousness. *Progress in Brain Research* **150** (2005) 81-98.
[12] Crick, F. *Věda hledá duši (Překvapivá domněnka).* Mladá Fronta 1997.
[13] Damasio, A.: Synchronous activation in multiple cortical regions: a mechanism for recall. *Seminars in Neuroscience* **2** (1990) 287-297.
[14] Davidson, D.: *Mental Events*. Reprinted in Essays on Actions and Events. Oxford: Clarendon Press 1970.
[15] Dennett, D.: *Consciousness explained.* Little, Brown 1991.
[16] Descartes, R. *Meditations on First Philosophy*. Preklad: Lafleur, L. J., NY: MacMillan 1985.
[17] DiCarlo, J. J.: Making faces in the brain. *Nature Neuroscience* 442 (2006) str. 664.
[18] Eguíluz, V. M. a spol.: Scale-free brain functional networks. *Physical Review Letters* **94** (2005) 018102.
[19] Engel, A. K., Singer, W.: Temporal binding and the neural correlates of sensory awareness. *Trends in Cognitive Science* **5** (2001) 16-25.
[20] Farkaš, I.: Recipročná kauzalita medzi mysľou a mozgom z neurovednej perspektívy. *Kognice a umělý život V*, Smolenice 2005, 117-124.
[21] Farkaš, I.: Samoorganizácia ako hybná sila dynamických vzorcov aktivity v mozgu a mysli. V zborníku *Kognice a umělý život VI*, Trešť 2006, 143-148.

[22] Fell, J. et al.: Human memory formation is accompanied by rhinal-hippocampal coupling and decoupling. *Nature Neuroscience* **4** (2001) 1259-1264.
[23] Frith, C., Perry, R., Lumer, E.: The neural correlates of conscious experience: an experimental framework. *Trends in Cognitive Sciences* **3** (1999) 105-114.
[24] Fries, P., Reynolds, J. H., Rorie, A. E., Desimone, R.: Modulation of oscillatory neuronal synchronization by selective visual attention. *Science* **291** (2001) 1560-1563.
[25] Fuchs, A. et al.: Spatiotemporal analysis of neuro-magnetic events underlying the emergence of coordi-native instabilities. *Neuroimage* **12** (2000) 71-84.
[26] Gerstner, W., Kistler, W. M.: *Spiking Neuron Models: Single Neurons, Populations, Plasticity.* Cambridge University Press 2002.
[27] Gödel, K.: Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme, I. *Monatshefte für Mathematik und Physik* **38** (1931) 173-98.
[28] Goldstein, J.: Emergence as a construct: History and issues. *Emergence: Complexity and Organization* **1** (1999) 49-72.
[29] Gray, C. M. et al.: Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature* **338** (1989) 334-337.
[30] Grossberg, S.: The link between brain learning, attention and consciousness. *Consciousness and Cognition* **8** (1999) 1-44.
[31] Haken, H.: *Synergetics.* Springer Verlag 1977.
[32] Hung, C. P., Kreiman, G. K., Poggio, T., DiCarlo, J. J. Fast readout of object identity from macaque inferior temporal cortex. *Science* **310** (2005) 863-866.
[33] Izhikevich, E. M., Desai, N. S.: Relating STDP to BCM. *Neural Computation* **15** (2003) 1511-1523.
[34] Izhikevich, E. M.: Spike-timing dynamics of neuronal groups. *Cerebral Cortex* **14** (2004) 933-944.
[35] Kelso, J.: Phase transitions and critical behavior in human bimanual coordination. *American Journal of Physiology* **15** (1984) 1000-1004.
[36] Kelso, J.: *Dynamic Patterns: The Self-Organization of Brain and Behavior.* MIT Press 1995.
[37] Kendler, K. S.: A psychiatric dialogue on the mind-body problem. *American Journal of Psychiatry* **158** (2001) 989-1000.
[38] Kim, J.: Problems in the philosophy of mind. In Honderich, T. (ed.): *Oxford Companion to Philosophy*. Oxford: Oxford University Press 1995a.
[39] Kim, J.: Review: Mental causation in Searle's biological naturalism. *Philosophy and phenomenological research* **55** (1995b) 189-194.
[40] Kim, J.: *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*. Cambridge MA: MIT Press 1998.
[41] König, P., Engel, A. K., Singer, W.: Integrator or coincidence detector? The role of the cortical neuron revisited. *Trends in Neurosciences* **19** (1996) 130-137.
[42] Lachaux, J-P. a spol.: Measuring phase synchrony in brain signals. *Human Brain Mapping* **8** (1999) 194-208.
[43] Le Van Quyen, M. et al.: Temporal patterns in human epileptic activity are modulated by perceptual discriminations. *Neuroreport* **8** (1997a) 1703-1710.
[44] Le Van Quyen, M. et al.: Unstable periodic orbits in human epileptic activity. *Physical Review E* **56** (1997b) 3401-3411.
[45] Legenstein, R., Naeger, C., Maass, W.: What can a neuron learn with spike-timing synaptic plasticity? *Neural Computation* **17** (2005) 2337-2382.
[46] Libet, B.: Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behavioral & Brain Sciences* **8** (1985) 529-566.

[47] Maxwell, N. The mind-body problem and explanatory dualism. *Philosophy* **75** (2000) 49-71.
[48] Penfield, W., Jasper, H.: *Epilepsy and the Functional Anatomy of the Human Brain.* Little, Brown & Co. 1954.
[49] Phillips, W., Singer, W.: In search of common foundations for cortical computation. *Behavioral and Brain Sciences* **20** (1997) 657-722.
[50] Popper, C., Eccles, J. C.: *The Self and Its Brain: An Argument for Interactionism.* Routledge 1984.
[51] Riehle, A. et al.: Spike Synchronization and rate modulation differentially involved in motor cortical function. *Science* **278** (1997) 1950-1953.
[52] Robb, D., Heil, J.: Mental causation. *Stanford Encyclopedia of Philosophy* 2003 http://plato.stanford.edu.
[53] Rodriguez, E. et al.: Perception's shadow: long-distance synchronization of human brain activity. *Nature* **397** (1999) 430-433.
[54] Roelfsema, P. et al.: Visuomotor integration is associated with zero time-lag synchronization among cortical areas. *Nature* **385** (1997) 157-161.
[55] Salinas, E., Sejnowski, T.: Impact of correlated synaptic input on output firing rate and variability in simple neuronal models. *Journal of Neuroscience* **15** (2000) 6193-6209.
[56] Schmid-Schonbein, C.: Improvement of seizure control by psychological methods in patients with intractable epilepsies. *Seizure* **7** (1998) 261-270.
[57] Searle, J. R.: Minds, brains and programs. *Behavioral and Brain Sciences* 3 (1980) 417-457.
[58] Searle, J. R.: *The Rediscovery of the Mind.* MIT Press 1992.
[59] Song, S., Miller,K., Abbott, L.: Competitive hebbian learning through spike-timing-dependent synaptic plasticity. *Nature Neuroscience* **3** (2000) 919-926.
[60] Sporns, O., Zwi, J. D.: The small world of the cerebral cortex. *Neuroinformatics* **2** (2004) 145-162.
[61] Steinmetz, P. E et al.: Attention modulates synchronized neuronal firing in primate somatosensory cortex. *Science* **404** (2000) 187-190.
[62] Stoljar, D. *Physicalism*. Stanford Encyclopedia of Philosophy. Stanford University 2001.
[63] Stopfer, M. et al.: Impaired odour discrimination on desynchronization of odour-encoding neural assemblies. *Nature* **390** (1997) 70-74.
[64] Strogatz, S. *Sync: The Emerging Science of Spontaneous Order*. NY: Hyperion Books 2003.
[65] Tononi, G., Edelman, G.: Consciousness and complexity. *Science* **282** (1998) 1846-1851.
[66] 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.
[67] Varela, F. et al.: The brainweb: phase synchronization and large-scale integration. *Nature Reviews Neuroscience* **2** (2001) 229-39.
[68] Varela, F., Thompson, E., Rosch, E.: *The Embodied Mind*. Cambridge, MA: MIT Press 1991.
[69] Vaas, R.: Problems of mental causation - Whether and how it can exist: A review of J. Kim's Mind in a Physical World, *Psyche* 8 2002. http://psyche.cs.monash.edu.au/v8/psyche-8-04-vaas.html
[70] Von der Malsburg, C.: Self-organization and the brain. In: *Handbook of Brain Theory and Neural Networks* (M. Arbib, ed.), MIT Press (2003), 2. edition, 1002-1005.
[71] Wegner, D. M.: *The Illusion of Conscious Will*. MIT Press 2003.
[72] Ziemke, T.: What's that thing called embodiment? *Proceedings of the 25th Annual Meeting of the Cognitive Science Society.* Lawrence Erlbaum 2003.
[73] http://en.wikipedia.org/wiki/Causality
[74] Burian, J.: Samoorganizace a kognice. V zborníku *Kognice a umělý život VI,* Třešť 2006, 107-118.

Tělo a mysl – východiska a alternativy

László Hajnal[1]

Abstrakt: Problém tělo-mysl vyvstává z intuitivního pochopení zásadní odlišnosti ontologického zakotvení zmíněných entit. Oddělení mysli a těla však implikuje nejméně dvě otázky. Za prvé, je-li mysl jiné přirozenosti než tělo, pak co je její podstatou? Za druhé, jestliže jsou tělo a mysl distinktní, jaká je – je li vůbec – příčinná následnost mezi činy a počitky mysli a těla? Jak se ovlivňují navzájem?

1 Východiska

Vztah těla a mysli je jedním ze zásadních problémů vědy i filozofie. Nakolik je tato otázka relevantní lze ukázat na příkladu anglosaské filozofie 20. století, která je bezprostředně i zprostředkovaně (např. přes kontexty lingvistické i filozofie jazyka) filozofií mysli nebo vědomí. Je důležité vzpomenout i tzv. naturalismus (scientismus?) – přesvědčení, podle kterého nás studium přírody přivede k odpovědím na všechny otázky, které vůbec lze položit. Pro antinaturalisty respektive antimaterialisty (tyto dva směry se liší) je mysl a její jednotlivé schopnosti (jazyk, otázky pravdy, rozum) zásadním a bezpečným východiskem jakéhokoli filozofického zkoumání, tvoří pevný bod, který nepotřebuje nebo neumožňuje další zkoumání sebe sama. Důsledný naturalista je naproti tomu nucen nahlédnout, že nemůže věřit ani sám sobě. Jelikož jsme součástí přírody, musíme si uvědomit, že veškerá naše kognitivní činnost (poznávání, reflexe) není řešením nýbrž součástí problému. Mysl, jestli vůbec existuje, sama čeká na vysvětlení podle stejných zásad, kterými vysvětlujeme veškeré další zákonitosti přírody. Jinými slovy: jestliže bude vysvětlení procesů mysli úspěšné, musí také objasnit, jak mysl vůbec poznává svět. Toto objasnění ale musí být popsáno pomocí stejné metody a zákonitostí jako všechny ostatní popisy přírody a přírodních zákonů.

V daném kontextu lze veškeré antinaturalistické ideje charakterizovat jako pokusy o popření strukturální a metodologické jednoty východisek a cílů kognitivních procesů (ať už je východisky a cíli myšleno cokoli). Pro úspěšného naturalistu je jakékoli vysvětlení myšlenkových procesů nutně důkazem kruhem. Důkazový kruh (a proto všechna důsledná naturalistická vysvětlení) se soustředí kolem jediného centrálního bodu – vztahu těla a mysli. Je však nezbytné uvést, že jistou cestou z důkazu kruhem je tzv. lokální naturalizmus; ten lze uplatnit konkrétně na cokoli, ale nikdy ne na všechno. Například mohu být naturalistou a uplatnit metody vědeckého zkoumání na veškeré cukrářské výrobky aniž bych se zabýval procesy mysli daného cukráře (a tím byl nucen pokládat si nepříjemné otázky ve vztahu k mé vlastní mysli).

V čem tedy spočívá problém vztahu těla a mysli? Na relaci tělo – mysl je možno se tázat mnoha způsoby. Zkoumaní by tak mělo začít pokládáním dobře definovaných otázek. Jestliže postulujeme základní otázky odpovídajícím způsobem, některé odpovědi budou předem vyloučeny, jiné pak upřednostněny. Jedním z mála uplatnitelných

[1] Katedra filosofie a dějin přírodních věd UK PřF, Viničná 7, Praha 2, E-mail: hajnall@cuni.cz

východisek je fakt, že v okcidentálním myšlenkovém okruhu je v dualitě tělo – mysl vždy přítomno jakési napětí, což může naznačit dobrý základ k dalšímu tázání.

V první řadě je třeba se ptát na podstatu fyzických a mentálních procesů (fyzickými procesy je myšlena např. neurální aktivita mozku, mentálními pak myšlení, poznávání, představivost...). Dle jednoho z možných náhledů jsou fyzické a mentální procesy z hlediska zevních projevů odlišné:

1. tělo je ze své podstaty fyzikální
2. mysl jakýmsi zvláštním způsobem fyzikální není
3. mysl a tělo jsou ve vzájemném vztahu a působení
4. fyzikální a ne–fyzikální nemůže být (dle klasických vědeckých teorií) v žádném vzájemném vztahu

Zdánlivě jen málo se liší následující způsob vyjádření:

Fyzikální procesy jsou a) dobře uchopitelné (zkoumatelné, kvantifikovatelné), b) objektivní (tím je myšleno, že se k výsledku se může dopracovat kdokoli, kdo použije zvolenou metodu), c) bez významového obsahu (tj. chovají se dle přírodních zákonů) d) ne-vědomé. Naproti tomu jsou mentální procesy a) neuchopitelné (klasickými metodami vědy), b) subjektivní (vázané k prožitkům a historii jedince), c) významuplné (intencionální), d) mohou být vědomé.

Rozdíl mezi uvedenými způsoby tázání souvisí s akcentováním nebo vypouštěním důrazu na svět významů a prožitků; v souladu se zvolenou metodou bude první typ zkoumání soustředěn na dvě různé substance, resp. spolupráci a vzájemný vztah fyzických a mentálních projevů a procesů. Druhý typ tázání se musí soustředit zejména na vědomé a intencionální jevy a jejich umístění v materiálním světě. Učiňme však myšlenkový pokus: v rozporu s výše uvedeným předpokládejme, že první přiblížení a hrubý náčrt problému tělo – mysl lze uskutečnit aniž bychom upřednostnili jeden nebo druhý způsob zkoumání. Uvedený způsob přiblížení předpokládá (nesmyslnou, leč ve vědeckém světě zcela běžnou) aproximaci: rozbor a řešení problému je nezávislé na historii, prožitcích a stanoviscích výzkumníka. Smysluplný pokus o takto sjednocený pohled vyžaduje nejdříve krátký historický přehled návrhů řešení problému tělo – mysl.

Na prvním místě je vhodné vzpomenout různé druhy dualismu. Dualistické hledisko předpokládá úplné nebo alespoň značné oddělení mentálního a fyzického světa. Pro dualistu je svět složen ze dvou substancí, které (v závislosti na typu dualismu) na sebe působí nebo nepůsobí; ve druhém případě se dostáváme k různým druhům paralelizmu, který se vyznačuje nekauzálním, neinteraktivním vztahem mezi mentálními a fyzickými událostmi. Systematický soulad mezi správnou neinteraktivní součinností těla a mysli řeší např. teorie okasionalismu nebo teorie (předstvořené) harmonie. Dle okasionalistického pojetí existuje jakýsi agent mimo náš svět (třeba Bůh, nebo bůh), který při každé události jednoho typu stvoří odpovídající událost typu druhého. Teorie harmonie předpokládá automatický soulad obou složek na základě dobře zkonstruovaného světa, tedy to, že v rámci obou oddělených světů odlišných substancí se odehrávají události právě tak, aby byl udržen soulad mezi mentálním a fyzickým (proto pociťujeme při doteku horké plotny bolest, i když horko není příčinou bolesti).

Představitelé interakcionalistického dualismu mají za to, že mentální procesy jsou příčinou fyzických projevů (a vice versa). Jejich hlavním problémem je vedle pojmoslovných otázek (jestliže jsou jednou substance dvě, pak jsou dvě, a když přesto

k interakci dochází, není ve skutečnosti substance pouze jedna?) riziko nutnosti rezignace na empirii. Zatím totiž nebyla objevena spojnice mezi oběma světy (Descartes předpokládal danou funkci u šišinky mozkové, Popper a Eccles podezřívali různá nastavení neurální sítě [7], Penrose a Hameroff zase předpokládali kvantové jevy v mikrotubulech [10, 21], abychom zmínili jenom několik teorií z mnoha).

Protipólem rozmanitých dualistických teorií je eliminativní (reduktivní) materialismus, hlásající odvoditelnost každého zdánlivého mentálního aktu od neurální mozkové aktivity, tj. neexistenci mentálních jevů jakožto entit per se. Eliminativní materialista tvrdí, že veškerá naše slovní zásoba, týkající se mysli a mentálních jevů se stane s postupujícím pokrokem v oblasti neurověd zbytečnou – nahradí ji přesná vědecká terminologie.

Je-li zaveden pojem reduktivního materialismu, zajisté počítáme i s existencí materialismu nereduktivního. Uvedené myšlenkové postupy mají většinou formu různých teorií identity, čímž se rozumí teze stejnosti mentálního a fyzického aktu. Teorie typových identit prohlašuje, že stejný typ mentálních aktivit znamená také stejný typ aktivit fyzických (i u různých osob), skromnější teorie exemplárních identit tvrdí, že každé konkrétní mentální aktivitě odpovídá nějaká aktivita fyzická. Jako příklad teorií druhého typu se často uvádí emergenční materialismus, tvrdící, že jedinečné kombinace fyzických událostí (událostí materiálního světa) vyvolávají mentální akty takovým způsobem, jejž nelze odvodit od jednotlivých fyzických aktů. Jistou podobnost s řečeným vykazují i supervenienční (nástavbové) teorie, snažící se prokázat mezi oběma typy aktů různé druhy složitých paralelních kauzálních závislostí [6, 12].

Vzpomeňme některé další názory: epifenomenalismus (fyzikální děje jsou příčinou mentálních aktů, ne však opačně), teorie duálního aspektu (každý projev má mentální i fyzický aspekt – i kámen má duši, jenom malou), logický behaviorismus (mentální akty jsou pouze pojmy, jejichž skutečným obsahem je výstup – dispozice k určitému jednání na základě vstupu – podráždění, signálu). Pro ilustraci si můžeme ještě připomenout evoluční naturalismus, který se snaží v biologickém kontextu boje o přežití dát do souvislosti mentální akty s adaptabilitou [3, 12, 18].

Jedním ze zajímavých (a pro další výklad klíčových) filosofických směrů je funkcionalismus, podle kterého fyzická realizace funkčního agenta nepatří k jeho podstatě. To, co činí funkční jednotku „správně fungujícím“ agentem, je způsob, jakým během svého fungování spojuje vstupy a výstupy. Dle funkcionalismu není mysl fyzická (materiální) ve stejném smyslu jako není materiální počítačový software, i když obojí může platně existovat jenom ve fyzické reprezentaci (software bez realizace, ve formě neinterpretovaného kódu je pouze potencialitou, nikoli programem). Jakýkoli mentální akt můžou vyvolat i úplně odlišné fyzické děje, do momentu, než splňují danou specifikaci vstup – výstup.

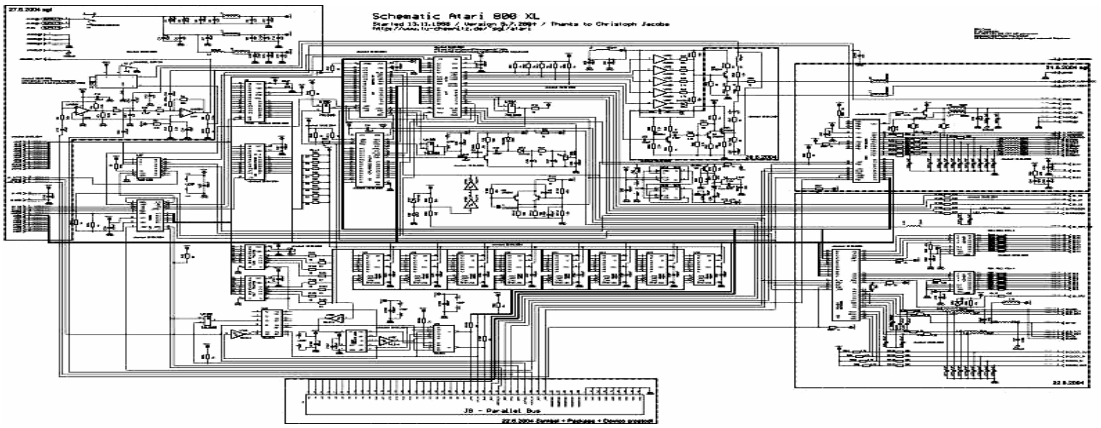
Funkcionalismus je zajímavý z více hledisek: je zodpovědný za rozšíření počítačové metafory mysli v kognitivních vědách. Dále: skrze Church – Turingovu hypotézu spojuje problémy mysli s otázkou hranic obecných teoretických struktur. Zjednodušeně řečeno, Church – Turingova hypotéza tvrdí, že všechno, co je precizně (symbolicky) vyjádřitelné v kterémkoli dobře definovaném jazyce, lze algoritmizovat a přeložit do jazyka počítačového [30]. Jelikož se obecně předpokládá, že veškeré vědění o světe je zaznamenatelné ve formě symbolického jazyka, nebylo daleko k hrubé, leč rozšířené aproximaci: celý svět je velký, dříve nebo později simulovatelný počítač. (Užití

termínu „počítač“ obvykle evokuje představu sekvenčního počítače von Neumannovy architektury, např. PC, kde je možné striktně oddělit hardware od software; kde na stejném počítači – hardware může být vykonáváno nepřeberné množství naprosto rozdílných programů – softwarů. Pro tyto počítače existuje striktní dichotomie mezi strojem a programem – tj. mezi hardwarem a softwarem. Paradigma mysli jako počítače implikuje mnohdy představu, že je možné oddělit mozek (tělo) od mysli, jako dva „nezávislé“ fenomény, kde tělo hraje úlohu hardwaru, zatímco mysl je software vykonávaný na hardwaru – mozku. V této souvislosti se mluví o sociokulturním zdroji lidské mysli, kde je mozek pouze irelevantním „hardwarovým vehiklem“ mysli [22]. Podstatně užitečnějším modelem je představa masivně paralelního neurálního počítače; tento návrh bude v textu rozveden později).

Klíčový význam funkcionalismu spočívá i v kontinuitě vědeckého poznání. Svou metodologií a propojením s analytickou filosofií budí zdání „bezpečné“ vědy – tedy vědy objektivní, vyrůstající z nejlepších tradic naturalismu. V kontinuitě vědeckého poznání spočívá i pramen pojetí vědy jako funkcionalismu (ve slabém smyslu). Například fyzik, zabývající se hydrodynamikou nejeví zájem o specifické, charakteristické vlastnosti jednotlivých látek (neboli o fyzické projevy), v posledku ho nezajímá ani fakt, je-li zkoumaná látka kapalinou. V tom je jeho přístup funkcionalistický: jako kapalinu specifikuje tu látku, na kterou lze aplikovat rovnice a definice hydrodynamiky, tak třeba bude jíl, sklo nebo krystal kovu posuzován za jistých podmínek jako kapalina. Uvedený příklad pouze ilustruje dávno známý paradox vědy: věda je materialistická jenom v empirickém pojetí, jakmile se však začne zabývat teorií, stává se idealistickou (přesněji řečeno mírně solipsistickou – Carnapův metodologický solipsismus [5]). Nechci tím říci nic víc než to, že vědecké teorie jsou v jistém smyslu slepé a necitlivé k pozorovaným „skutečnostem“ a zkušenostem vnějšího světa (již Jerry Fodor upozorňuje na fakt, že kognitivní teorie jsou zhusta necitlivé k tomu, jsou-li vstupy mysli pravdivé, nebo nikoli – i když v dalším si ukážeme, že takto položená otázka je nepodstatná [12]). Tématu se dotýká i problém realismu: kdyby teoretické termíny nebyly prázdné, problém slepoty vědy k světu by neexistoval.

Postulát, že věda je funkcionalistická ve slabém smyslu je ilustrovatelný na následujícím příkladu: klasické pojetí vědy uvažuje o vědeckých popisech (například o rovnicích výbuchu atomové pumy) jako o modelech, ve kterých se děje „to samé“, jako při skutečném výbuchu (alespoň co se týče podstaty) – tady je funkcionalistický rys. Funkcionální identita je aplikovatelná i na počítačovou simulaci (neboli na zopakování „skutečného“ výbuchu v prostředí počítače), což jenom umocňuje pocit oprávněnosti existence funkcionalistického pojetí vědy. Ale – při doběhnutí programu simulace výbuchu počítač přece nevybouchne! Naproti tomu kognitivní funkcionalismus věří, že počítačový algoritmus, popisující akty mysli sám tvoří mysl a proto může nabýt vědomí, mít prožitky atd. – jedním slovem, po simulaci rovnic dojde k „výbuchu“ ve vyšší hierarchické hladině. Pro kognitivní funkcionalismus (který je vskutku silným funkcionalismem) je příznačná teze architektonického pojetí realizace mysli. V nejzazší podobě je vyslovena H. Simonem a A. Newellem (oblast výzkumu AI): „... provozovat symbolický systém vyžaduje stejné schopnosti, jako provozovat mysl.“ [29] Jinak řečeno, mysl je symbolickým systémem a naopak – správně fungující symbolický systém lze popsat jako mysl. Hypotéza symbolické mysli je ústřední myšlenkou klasické kognitivní vědy, která se otevřeně hlásí k metodologickému solipsismu a k individualismu.

I když kognitivní funkcionalismus má mnoho slabých míst, pro svou univerzální teoretickou základnu je stále populární, protože je-li každý proces komputací, pak procesy mysli jsou také komputační; teoretický rámec je tedy dán a stačí najít odpovídající algoritmus (i když nezodpovězeným problémem zůstává otázka po původu komputačního zařízení). Abychom se mohli přesvědčit o nedostatečnosti myšlenkového rámce kognitivního funkcionalismu a dobrat se jiných variant, je nutné přiblížit si ve zkratce některá teoretická východiska alternativních přístupů ke zkoumání vztahu tělo – mysl.



Obrázek 1. Schéma hardware – počítače Atari 800 XL

2 Alternativy

2.1 Modely a teoretická východiska

V posledních desetiletích se jak v přírodních, tak i v humanitních vědách stal mimořádně populárním směr hlásající, že matematika není vhodná pouze pro popis kvantitativních nýbrž i kvalitativních jevů [6, 17, 28]. Jednou ze základních otázek se pak stává: do jakých kvalitativně odlišných stavů se může systém dostat? Na podobné otázky lze odpovídat do jisté míry pomocí teorie chaosu; termodynamika zde dostala zcela novou podobu studiem nerovnovážných systémů. Ukazuje se, že druhá věta termodynamiky nevede pouze k rovnovážným stavům, ale že se během cesty k rovnováze objevují složité struktury a časové cykly. Teorie chaosu se zabývá takovými jevy, jejichž významná část je ze své podstaty deterministická a průvodní podmínky stochastické – ve své podstatě je tedy naukou o nespojitých jevech.

Jak se tedy teorie chaosu má k termodynamice? V období mezi Fourierem a Gibbsem bylo v podstatě jasné, že termodynamika se zabývá ireversibilními makroskopickými ději probíhajícími v konečném čase. Gibbs – i když zachoval název termodynamika – věděl, že se zabývá naukou o rovnováze. Vědecký aparát druhé poloviny 20. století – teorie chaosu, synergetika, teorie disipativních struktur – klade důraz na klasifikaci konečných stavů dějů, probíhajících v termodynamických systémech. Může-li mít systém několik kvalitativně rozdílných konečných stavů, je důležitá otázka, jak lze přestoupit z jednoho stavu do jiného. Názorně můžeme říci, že zkoumáme-li několik ve svých parametrech mírně odlišných systémů, můžeme v jistých případech nalézt bifurkační body, ve kterých se dané systémy začnou od sebe zásadně odlišovat. Takovéto

systémy jsou strukturálně stabilní. Studium dynamických systémů, které jsme získali zobecněním průsečíku deterministických a stochastických modelů, nás může přivést k možnosti určení (rozpoznání) bifurkačních bodů nebo hranic stabilních průběhových drah libovolného otevřeného nelineárního systému. Dostatečně propracované průsečíkové modely mohou popisovat chování systémů s vysokou přesností. Fyzikální systém je ve své podstatě analogový počítač, který je naprogramován tak, aby simuloval své vlastní chování. Ale matematika nedospěla k teorii chaosu pouze prostřednictvím fyziky, ale také studiem fraktální geometrie a hledání řešení nelineárních diferenciálních rovnic, které nelze řešit explicitně, nýbrž pouze pomocí matematických modelů. Pomocí aparátu výše zmíněných teorií se pak dostáváme blíže k možnosti lepšího popisu některých netermodynamických systémů s termodynamickým základem (NESYTEZ).

2.1.1 Netermodynamické systémy s termodynamickým základem (NESYTEZ).

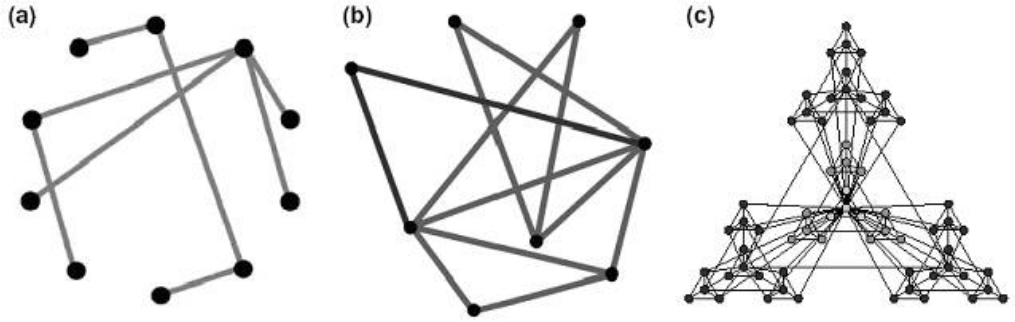
Abychom mohli přistoupit k vysvětlení fenoménu NESYTEZ je nutné si nejdříve něco říci o pojmu „entropie“. Pojetí entropie bylo definováno ve fenomenologické, respektive statistické termodynamice. Vynořuje se ale otázka, zachová-li si tento v termodynamice bezpochyby zásadní pojem svou důležitost i při zkoumání takových systémů, které jsou sice na nižších hierarchických úrovních popisu systémy termodynamickými ale na úrovních vyšších nikoli. Za takovéto systémy lze považovat biologické objekty nebo také průmyslová chemická zařízení. Pro osvětlení otázky je nutno si uvědomit, že a) tok materiálu a energie je doprovázen významným tokem informací b) základní termodynamické procesy se spojují do dějů vyššího řádu.

Dále je potřebné si uvědomit rozdíly mezi fyzikálním (termodynamickým) a matematickým (informatickým) pojetím entropie. Entropie jako informatický pojem byla definován Shannonem a je dnes považována za jeden ze základních pojmů pravděpodobnostního rozložení. Shannonovská entropie je analogická Gibbsovu pojetí. Zhusta diskutovanou, dodnes nevyjasněnou otázkou zůstává, je-li pojetí entropie v termodynamice a informatice obdobné pouze na úrovni matematické abstrakce (na formální úrovni) nebo je-li mezi těmito pojmy hlubší obsahová souvislost. Podle neklasického termodynamického výkladu pojmu entropie (Gibbs, Szilárd, Brillouin) provádíme-li měření libovolného systému, klesá množství termodynamické entropie o množství získané informace. Dle mého názoru je předpoklad negativní semidefinice sumy termodynamické a informatické entropie nesprávně položenou otázkou. Předpokládáme totiž, že se jedná o dvě nezávislé kategorie. Ve skutečnosti neexistuje nic takového jako tok informace nezávislý na toku energie a látky. Tok informace je speciálním případem energetických toků, který se však neřídí všemi zákony termodynamiky.

Pojem entropie je tedy používán v různých vědních oborech (biologie, ekonomie, lingvistika atd.). Musíme se však podívat jak se projevuje entropie v NESYTEZ. Asi nejdůležitějším příkladem NESYTEZ jsou biologické struktury. Vyšší úrovně hierarchie NESYTEZ jsou sice do jisté míry ohraničeny základními axiomy termodynamiky, ale na těchto úrovních již vznikají jiné typy interakcí, objevují se emergentní jevy a inherentně nepředvídatelné události a procesy. Kumulativní efekty v systémech často vedou ke vzniku paměti (ukládání informace), kdy systém zaznamenává svou historii v tom smyslu, že jeho současné vlastnosti závisejí na jeho předchozích vlastnostech. Pokud vnější parametry mají po určitém cyklu znovu stejné hodnoty, pak na rozdíl od jednoduchého

systému, který bude mít zcela stejné chování jako v cyklu předchozím, systém s pamětí se chová odlišně v závislosti na předchozím stavu nebo stavech. Matematicky lze takový dynamický systém popsat jako markovovský proces s diskrétním časem. Biologické systémy provádí neustávající látkovou, energetickou i informační výměnu se svým okolím. Jsou vysoce uspořádané a tak jejich entropie nedosahuje maxima. Některé jevy je možno vykládat jako stabilní stacionární procesy (ve velice specifických případech podléhají zákonu minimální produkce entropie), další jevy jsou cyklické v prostoru a/nebo v čase, nebo mají více stacionárních konečných stavů.

Komplikaci tvoří i synchronní povaha jevů v biologických systémech, jejich vzájemná provázanost a zvláštní způsob propojení. Jedním z perspektivních pohledů na způsob propojení nexů NESYTEZ je pattern sítí malého světa (je-li síť hierarchická, přísně pravidelně propojena, vzdálenosti od bodu k bodu mohou být obrovské, jelikož je - v případě lokálně propojených sítí - nutné překonávat je ergodicky. Je-li síť úplně náhodná, vzdálenosti jsou častokrát menší, ale současně je nízký i klasterizační koeficient sítě. Utvoříme-li však dobře strukturovanou pravidelnou síť a přidáme několik málo náhodných propojení (Granovetterových mostů), dostaneme fascinující útvar: síť malého světa – small world network. V těchto sítích jsou vzdálenosti mezi prvky o několik řádů nižší, nežli v lokálně propojených sítích; viz obr. 2).



Obrázek 2. Tři typy sítí: a) náhodná, b) síť malého světa c) síť hierarchická

2.2 Mysl jako zvláštní případ NESYTEZ

Předpokládáme-li, že mysl a tělo jsou nějakým způsobem propojeny, otázku důležitosti těla nelze opominout. Již John Searle zdůrazňoval význam wetware – jakéhosi biologického hardware se širokou paletou zvláštních a neopakovatelných vlastností [24]. Pokusme se tedy sloučit výše uvedená východiska a předpokládejme, že to, co nazýváme myslí, je produktem vzájemného působení velkého množství různorodých, aktivních, samostatných jednotek (komponentů), spojených do sítě malého světa. Předpokládejme dále, že jednotlivé komponenty sítě lze považovat za Kauffmanovy autonomní agenty, které se starají o sebe a mohou se stát (nutně jsou?) předmětem evoluce darwinovského typu. Užitečnou představu dostaneme, použijeme-li model založený na konekcionistickém pojetí mozku i mysli. Moderní hypotézy tvrdí, že lidský mozek je utvořen z neuronů navzájem propojených pomocí synaptických spojů; ty v průběhu učení neustále vznikají a časem zanikají (podílejí se tak na neobyčejné plasticitě mozku). Architektura mozku je tedy určená spoji mezi neurony, jejich inhibičním nebo excitačním charakterem a také jejich intenzitou. Lze předpokládat, že schopnost lidského mozku vykonávat kognitivní

aktivity a současně sloužit i jako paměť je ve vysoké míře dané jeho architekturou. Na základě uvedeného se mimořádně vhodným ukazuje přístup k mozku/mysli jako k masivně paralelnímu distribuovanému počítači, obsahujícímu mnoho miliard neuronů, jejich shluků i vyšších komplexů, které jsou mezi sebou propojeny formou sítí malého světa. „Program“ takovéhoto paralelního počítače je (do jisté míry a v souladu s teorií emergentních systémů) přímo zabudován do architektury sítě. Jestliže budeme považovat lidský mozek (a v širším pojetí celé tělo) za takovýto paralelní počítač, pak dojdeme k závěru, že jej nelze přeprogramovat bez změny jeho architektury a – co je důležitější – změna architektury povede (v souladu s předpoklady evoluční biologie) i ke změnám v programu.

Abychom mohli dále rozvinout hypotézu jednoty těla(mozku)/mysli, je nutné si přiblížit pojmy jako autonomní agent, distribuovaná síť či emergentní systém. Autonomní agent v Kauffmanově pojetí je „ ... bytost či systém, jehož aktivita není přísně příčinná, ale bere ohled na své minulé a současné vnitřní stavy, a také na stavy okolí. Na podněty prostředí agent neodpovídá podle „slepých“ zákonů akce a reakce, ale koná podle toho, jak vyhodnotil situaci. ... Agentem je jakákoli živá bytost, agentem je buňka, agentem jsou i jisté produkty umělé inteligence, umělého života nebo pokročilé robotiky.“ [14] Agent jako jednotka nemusí být nijak zvlášť chytrý, stačí mu několik málo pravidel, trochu paměti, možnost rozhodování a svoboda upravit si vnitřní reprezentaci světa na základě minulých zkušeností. Pokud tedy autonomní agenty navrhneme tak, aby měly paměť, získají schopnost znát údaje o prostředí. Je-li pak agent schopen odhadovat trendy vývoje, dokáže se na jistá místa vracet s jistými očekáváními.

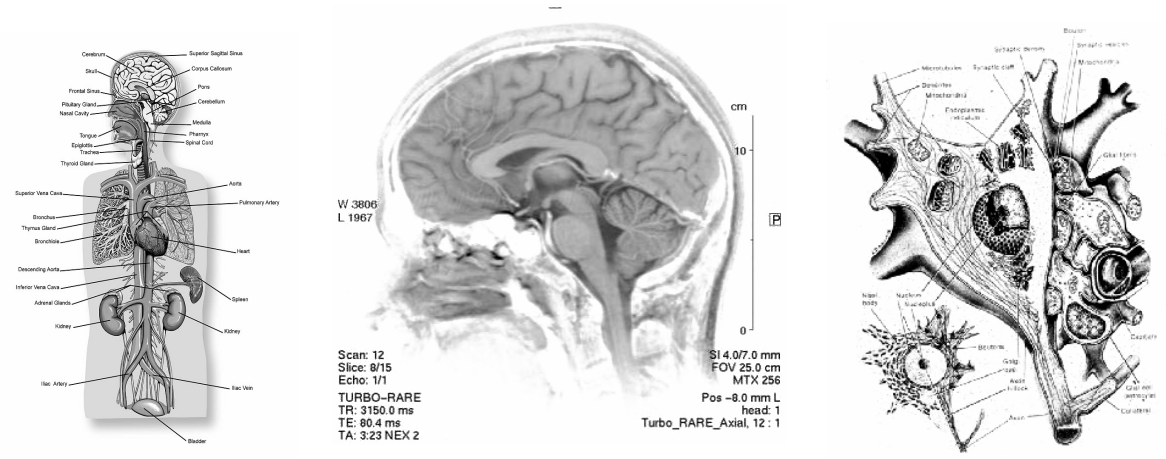
Zajímavé je, že ve všech distribuovaných systémech si po čase jisté agenty opatří mylné názory. Začnou se chovat v kontextu svého prostředí nepatřičně, ať už je to kvůli konfliktu motivací, anebo z jiného důvodu. Prostředí se změnilo, ale ony jako by to nevěděly. Opakují zastaralé vzorce, uvízly v minulosti. Jejich chování už neodráží aktuální realitu. V evolučních programech takové agenty vymřou, nemají potomky, nebo jsou zatlačeny na periferii. Celý systém je však ještě o něco zajímavější. Můžeme vytvořit počítačový model prostředí s mnoha agenty, které interagují, aby dosáhly cíle. Komplexu agentů můžeme zadat složitý úkol, který by jednotlivé agenty nezvládly, ale jejich interagující síť si se zadáním poradí. Dostáváme se do oblasti distribuovaného zpracování dat, což znamená, že rozdělíme práci mezi několik procesorů, anebo síť agentů. Jedním z účinných způsobů je vytvořit početnou populaci poměrně hloupých agentů, které na dosažení cíle pracují společně – stejně jako společně pracuje s určitým cílem mraveniště. Distribuovaný systém je nutno programovat zdola nahoru. Program definuje chování jednotlivých agentů pouze na nejnižší strukturální úrovni, chování celého systému však nijak definováno není; místo toho se objevuje jako výsledek nesčetných nenápadných interakcí na nejnižší úrovni. V systému pak vznikají emergentní jevy – vynořující se chování, které sice není naprogramované, ale je důsledkem programování [20].

Pro vynořující se skupinové chování nelze určit pravidla; zásady lze určit pouze pro jednotlivé agenty. Lze je naprogramovat, aby za určitých okolností kooperovaly a za jiných podmínek soupeřily. Je možné je instruovat, aby za úkoly šly tvrdohlavě a o samotě, anebo aby současně pomáhaly jiným agentům. Ale výsledek těchto interakcí se naprogramovat nedá; emergentní chování je kolísavé, nepředvídatelné. V počítačových modelech se proto používají motivátory – programové koeficienty, které podporují cíle činnosti. Jsou potřeba, protože do sítě spojené agenty se dovedou učit, a to i takovým

způsobem, který je od cíle odvede. Je nutno najít způsob, jak uložit původní cíl, aby se neztratil [29]. A je současně nutné určit, které parametry vnějšího světa jsou důležité, které se vyplatí sledovat a které nikoli; agent, který neví, co má měřit a zaznamenávat, bude nutně neúspěšný. Bakterii třeba vůbec nemusí zajímat barva roztoku, ale obsah cukru v něm je už mnohem důležitějším parametrem [14].

Za agenty můžeme považovat nejenom základní stavební jednotky – které se v reálném světě stejně určují mimořádně obtížně, jde-li to vůbec – ale i jejich organizované celky, skupiny či komplexy. Tak agentem může být buňka, játra, člověk ale i sídliště, město nebo stát. Záleží pouze na tom, jaké si zvolíme měřítko – většinou v každém rozlišení najdeme něco, co lze nazvat na základě původní definice agentem. Louka není o nic méně agentem než včela, pouze my máme problém vymezit hranice louky; hranice včely více-méně známy jsou.

Dostáváme se tedy k zásadnímu bodu: vysoce koordinovaná komplexní činnost nemusí být výsledkem komplexní inteligence, leč může vzniknout pomocí samoorganizace a vynořovat se na základě složitého propletence nesčetných interakcí, paměti a učení [11]. Lze tedy považovat neurony (resp. jejich shluky) za agenty, neurální komplexy za agenty vyššího organizačního řádu a jejich souhrn (mozek, nervovou síť, tělo) za jakéhosi makroagenta, fungujícího jako nonhierarchická distribuovaná síť malého světa? Jestliže ano, pak by mysl nemusela být substancí oddělenou od těla, nýbrž projevem jistého způsobu fungování wetware v souladu s poznatky teorie chaosu, sítí, emergence atd. Po sobě následující zpětněvazebné události by mohly indukovat změnu funkce jednotlivých propojených systémů (stejně jako např. v případě proteinů) a vyvolat odlišný způsob propojení, vyhovující nové situaci. *Pojetí* mysli jako projevu těla by nám umožnilo odloučit se od zjevného i skrytého dualismu, stejně jako od krajných pojetí (pouze mysl – žádná mysl). V předkládaném konceptu tedy tvoří mysl a tělo integrální celek: jsou to pouze dva pohledy na stejný objekt. Mysl tak získává další rozměr – analogovou tělesnost, celé spektrum barev, které komputační modely vyloučily – a může se vrátit ze světa idejí do světa forem.



Obrázek 3. Tři úrovně wetware: lidské tělo, mozek a neurální buňka

Výběrová bibliografie:

[1] Barabási, A. L.: Behálózva – A hálózatok új tudománya. Budapest: Magyar Könyvklub 2003.
[2] Bednář, M., Fraňková, V., Schindler, J., Souček, A., Vávra, J.: *Lékařská mikrobiologie.* Praha: Marvil 1996.
[3] Blass, T.: *A Life of Inventiveness an Controversy*; In.: Portraits of Pioneers Psychology. Vol. 11. Washington: Amer. Psychological Association 1996.
[4] Buchanan, M.: *Nexus: Small Worlds and the Groundbreaking Science of Networks*. New York: W.W. Norton & Company 2002.
[5] Carnap, R.: *The Logical Structure of the World. Pseudoprobleme in Philosophy*. London: Berkeley 1967.
[6] Dennett, D. C.: *Consciousness Explained*. New York: Little Brown 1991.
[7] Eccles, J., Popper, K. R.: *The Self and its Brain*. New York: Springer 1977.
[8] Edelman, G. M.: *Neural Darwinism: The Theory of Neuronal Group Selection*. New York: Basic Books 1987.
[9] Granovetter, M.: The Strength of Weak Ties. *Americ Journal of Sociology* **78** (1973) 1360-1380.
[10] Hameroff, S. R.: Quantum Coherence in Microtubules: a neural basis for emergent consciousness? *Journal of Consciousness Studies* **1** (1994) 98 – 118.
[11] Holland, J. H.: *Hidden Order: How Adaptation Builds Complexity*. Cambridge: Perseus 1996.
[12] Kampis, G.: *Self-Modifying Systems in Biology and Cognitive Science*. Oxford: Pergamon 1991.
[13] Kauffmann, S.: *Čtvrtý zákon – cesty k obecné biologii*. Praha: Paseka 2004.
[14] Markoš, A., Kelemen, J.: *Berušky, andělé a stroje*. Praha: Dokořán 2004.
[15] Markoš, A.: *Povstávání živého tvaru*. Praha: Vesmír 1997.
[16] Markoš, A.: *Tajemství hladiny – hermeneutika živého*. Praha: Dokořán 2003.
[17] Milgram, S.: The Small-World Problem. *Psychology Today* **1** (1967) 60-67.
[18] Minsky, M.: *Society of Mind*. Cambridge, MA: MIT Press 1985.
[19] Mizruchi, M. S.: What Do Interlocks Do? *Annual Review of Sociology* **22** (1966) 270-298.
[20] Nishimura, S. I., Ikegami, T.: Emergence of Colective Strategies in a Prey-Predator Game Model. *Artificial Life* **4** (1997) 243-260.
[21] Penrose, R.: Shadows of the Mind. Oxford: Oxford University Press 1994.
[22] Putnam, H.: *Representation and Reality*. Cambridge, MA: MIT Press 1988.
[23] Raup, D. M.: *O zániku druhů*. Praha: Nakladatelství Lidové noviny 1995.
[24] Searle, J.: Minds, Brains and programs. *Behav. Brain Sci.* **3** (1980) 417-424.
[25] Smith, H. M.: Synchronous Flashing of Fireflies. *Science* **82** (1935) 151.
[26] Sonea, S., Panisset, M.: *A New Bacteriology*. Boston: Jones and Bartlett 1983.
[27] Watts, D. J.: *Small Worlds – The Dynamics of Networks between Order and Randomness.* Princeton: Princeton University Press 1999.
[28] Watts, J. D., Strogatz, H. S.: Collective Dynamics of „Small-World“ Networks. *Nature* **393** (1998) 440-442.
[29] Woolridge, M.: *Reasoning About Rational Agents*. Cambridge: MIT Press 2000.
[30] http://plato.stanford.edu/entries/church-turing/

Prožívání epizodických situací

Ivan M. Havel[1]

Abstrakt. Práce je pokusem o reflexivní analýzu lidského prožívání komplexních epizodických situací. Postupně je na příkladech (hlavně z beletrie) probíráno deset hlavních modalit prožívání (smysl pro čas, prostor, strukturu scény, zápletku, působení, pocit aktéra, vlastní tělesnost, vědomí sebe, smysl pro druhého, socialita), u většiny je navíc věnována pozornost prožitkům na horizontu subjektivní zkušenosti. V pozadí úvah je pokus najít styčná místa mezi dvěma odlišnými přístupy v kognitivní vědě, z nichž jeden je založen na tzv. perspektivě první osoby (typické pro fenomenologii) a druhý na perspektivě třetí osoby (typické pro vědu).[2]

1 Úvod

Začnu příkladem.[3] Je to první věta povídky E. A. Poa „Zánik domu Usherova“:

Příklad 1. (Jezdec)

> *Jel jsem koňmo, sám, po celý chmurný, tmavý, tichý podzimní den, kdy mraky visely tísnivě nízko na obloze, neobyčejně pustou krajinou a konečně s večerními stíny octl jsem se před melancholickým domem Usherů.* ([39], s. 7)

Tato věta možná naladí čtenáře na ponuře romantický příběh, který je třeba číst někde na samotě, v pohodlném křesle u lampy, nejlépe za tmavého, chmurného, tichého podzimního dne. Zde však nemáme než tu jednu větu – budiž nám příkladem vylíčení jedné konkrétní *epizodické situace*.[4] Podobné situace budou hlavním tématem této studie. Jde o elementární smysluplné segmenty lidské zkušenosti, pro něž je typické, že je lze z principu vyprávět (je *o čem* mluvit), a to v první osobě (je to situace *někoho*, kdo se v ní ocitá, prožívá ji), lze si je zpravidla pamatovat a vybavovat si je jako vzpomínky a v jejich osobitosti je lze navzájem rozlišovat a odděleně vyjímat z nediferencovaného pozadí. Mohou to být prožitky krátké (*ocitnutí se před domem Usherů*) i delší (*celodenní cesta pustou krajinou*).[5] Jak vidno, jedna taková situace může obsahovat jinou situaci jako svou část, záleží, k čemu se upíná náš zájem.

Z odstupu si lze představit, že celoživotní příběh[6] každého z nás je utkán z nepřeberného množství rozličných epizodických situací, které na sebe rozmanitě navazují, prolínají se a překrývají. Opakuji: jen z odstupu, což znamená v teoretické

---

[1] CTS (UK a AV ČR), Jilská 1, 110 00 Praha 1, Česká republika, E-mail: havel@cts.cuni.cz

[2] Tento text vznikl podstatným přepracováním a rozšířením práce [21] s použitím některých partií z [17] a z připravované práce [22]. Ostatní vlastní použité práce cituji na příslušných místech.

[3] Příklady jsou tištěny kurzívou k odlišení od (delších) citátů z odborné literatury, tištěné petitem.

[4] Důležitá poznámka: pokud v dalším textu píši „situace“ nebo „epizoda“, rozumím tím vždy „epizodická situace“.

[5] V této studii užívám převážně výraz „prožitek“ (a „prožívání“) pro typicky krátkou a souvislou zkušenost (a navíc s důrazem na to, že jde o zkušenost uvědoměnou) k odlišení od obecněji pojaté „zkušenosti“, která může být i dlouhodobá, opakovaná, případně i neuvědomovaná (sbírání zkušenosti, životní zkušenost apod.).

[6] Později jej budu nazývat „panoráma života“.

reflexi. Svůj vlastní život nikdy nevidím jako celek, budoucí situace neznám a většinu dříve prožitých situací si nepamatuji.

V běžném životě prožívám[7] každou aktuální situaci jako jedinečný, jednolitý a smysluplný celek, aniž bych své prožívání nějak tematizoval, tutéž situaci však mohu – pokud chci – *reflektovat* (podrobit *reflexi*) z jistého teoretického odstupu, zaměřit na ni pozornost, a to buď v její konkrétnosti, nebo jako na instanci obecnější, konceptuálně (pojmově) uchopitelné epizodické situace. V rámci takto chápané reflexe pak mohu i rozlišovat určité charakteristické *modality prožívání situací* [8] jako je například smysl pro čas, prostor, scénu, zápletku apod. (viz sekce 3). Přitom mohu vždy volit mezi dvěma epistemicky odlišnými tendencemi: první, subjektivizující tendence, je založena na *perspektivě první osoby* (hledisko prožívajícího „já“ – budu ji dále označovat zkratkou 1PP), druhá, objektivizující tendence, se naopak snaží důsledně držet *perspektivu třetí osoby* (pohled „odnikud“ či „odkudkoliv“, jak jej známe i z přírodních věd – dále 3PP).[9] Aby nedošlo k nedorozumění: při obou tendencích je řeč na stejné téma: jaká je daná situace *pro někoho* (totiž pro toho, kdo ji prožívá), nikoliv jaká je „sama o sobě“. V perspektivě první osoby situaci prožívám (jakoby) *já*, maje přitom „privilegovaný“ přístup k svým vlastním prožitkům; naproti tomu v perspektivě třetí osoby ji prožívá myšlený *subjekt situace* [10] (tedy „on“ – hlavní osoba situace, protagonista), o němž proto hovoříme v třetí osobě singuláru (nebo též v první osobě plurálu: *my* ve významu kdokoliv).[11] Žádnou z obou tendencí ovšem nelze dovést do „čisté“ podoby, volíme-li jednu, zpravidla se připlete i druhá. Pojmově je dovedeme rozlišit (jako pohled či popis v 1PP resp. 3PP), avšak v průběhu analýzy vždy jen mezi nimi oscilujeme, majíce je za dva (nedosažitelné) póly reflexe.

V této studii se pokouším tematizovat nejdůležitější modality prožívání epizodických situací, a to v obou zmíněných perspektivách (1PP a 3PP). Předložím k úvaze otázku, zda v případě hraničního (čili prahového) prožívání v té či oné modalitě nedochází ke sblížení (střetnutí, překrytí či splynutí) obou perspektiválních tendencí, což by mohlo vrhnout nové světlo na vzájemný vztah kognitivní vědy (jakožto vědy v třetí osobě) a

---

[7] Viz pozn. 11 níže.

[8] Z mnoha významů slova „modalita“ jsem zvolil ten, který připomíná „modality vnímání“ (zrak, sluch, hmat,…). Ty jsou v přirozeném prožitku navzájem zpravidla úzce provázány (příklad: zrak, sluch a hmat při hře na hudební nástroj), lze je však zkoumat i reflektovat odděleně. Totéž platí i v našem pojetí modalit prožívání (viz kap. 3).

[9] Termíny “perspektiva první osoby”a “perspektiva třetí osoby” [*first-person / third-person perspective*] jsou běžné v současné západní literatuře z kognitivní vědy a filosofie. Proto je zde užívám, i když v některých případech mohou být zavádějící. Pro vztah k druhému (přítomnému) subjektu si později zavedeme též perspektivu druhé osoby („ty“, 2PP).

[10] Takto budu i nadále užívat slovo „subjekt“ (ve shodě s literaturou v kognitivní vědě a tradiční psychologii).

[11] Pokud se teoretik hlásí ke druhé ze dvou zmíněných tendencí, pak se zpravidla snaží zamlčet své vlastní, zkoumající a teoretizující já (po způsobu vědy) – zda a do jaké míry je to vůbec možné, je jiná otázka. Od čtenáře žádám určitou vstřícnost při orientaci ve spleti různě míněných podmětů, kdy „já“ a „my“ může být jednou autorské (já, autor, a my, autor + čtenář, případně my, badatelé), jindy zástupné (já, příklad člověka, a my, všichni lidé). V anglické literatuře se často používá i sugestivní „you“ (ve smyslu: „čtenáři, platí to o tobě, stejně jako o každém“).

fenomenologie (založené na perspektivě první osoby).[12] Při výkladu se budu i nadále opírat o příklady epizodických situací převzaté (kromě dvou) z literární fikce.[13]

2 Vzpomínky, sny a fikční situace

Připomeňme si epizodu jezdce na koni z povídky E. A. Poa (Příklad 1). Jediná věta, i když vytržena z kontextu, nám nabídla příklad epizodické situace. Snad by takovéto vyjímání z kontextu mohlo znervóznit literárního teoretika, který musí brát do úvahy celek literárního díla a další reálie, zde je tomu jinak: určitá „izolace“ konkrétní situace je pro naše účely i výhodná – nejde nám o svět ani lidský život v jeho úplnosti.

Příklady z beletrie nejsou jediné, které lze snadno vyjmout ze souvislostí a zabývat se jimi samými. Použitelné jsou v tomto směru samozřejmě i vzpomínky na minulé epizody z vlastního života, ba i (a to zajímavě) i vlastní sny. Jejich výhodou je, že je máme v jistém smyslu „z první ruky“. Následuje příklad mé vlastní vzpomínky na reálný zážitek:

Příklad 2. (Kalifornie)

> *Stojím nad pobřežními útesy severní Kalifornie a dlouho, nejméně hodinu napjatě sleduji, jak se rozběsněné vlny Pacifiku s úpornou vytrvalostí opakovaně vrhají na masivy skalního kotle pode mnou a jak se po každém výpadu, zpěněny nezdarem, zas znovu rozhoupávají k novému, pokaždé jinak mířenému útoku.*

Pamatuji si to docela dobře, byl to zážitek spíše estetický, nicméně – s použitím metaforického jazyka – jej vyprávět jak vidno lze. Co je však důležité: odehrálo se to před 17 lety, mezitím jsem zcela zapomněl, co se dálo předtím a co potom, ba i kde to přesně bylo. Takže ve vzpomínce je to epizodická situace zcela samostatná, bez vnějších souvislostí a vztahů.

Je to ovšem vzpomínka a je vhodné si udělat malou odbočku o *paměti*. V psychologii a neurovědách je běžné rozlišovat různé typy paměti podle modu, funkce a obsahu a toto rozlišení lze mnohdy podpořit i určením specifické části mozku, která má ten či onen typ paměti na starosti ([28], kap. 3). Především se rozlišuje implicitní, *procedurální paměť* (např. pro dovednosti) a explicitní, *deklarativní paměť* (např. pro fakta); tato druhá se dále dělí na *sémantickou paměť* (encyklopedické či jinak obecné poznatky a pojmy) a *epizodickou paměť* (pro vlastní prožité události, příběhy a situace) [48, 49]. Tento poslední typ paměti nás bude zajímat.

Zatímco sémantická paměť se více či méně týká společného světa lidí a v té míře je i kulturně sdílena, náleží epizodická (autobiografická) paměť konkrétnímu jednotlivci a vztahuje se na jeho vlastní životní zážitky a zkušenosti. Toto rozlišení má jistě svou

---

[12] V této práci se nebudu zabývat fenomenologií v klasickém smyslu (rozvinutou Husserlem, Heideggerem a dalšími) a tudíž ani strukturou předreflexivního vědomí, ani obecnými podmínkami možnosti reflexivní analýzy – tu budu prostě předpokládat. Dodávám, že velmi zasvěcenou analýzu vztahu mezi (1) klasickou fenomenologií, (2) „fenomenologií“ v terminologii analytické filosofie a (3) kognitivní vědou podává Zahavi v knize [57] (srov. též [11]).

[13] Jsou výstižné a inspirativní – přinejmenším nejsou nudné ani banální. V zájmu výkladu si budeme hrát na skutečnost – každý si ostatně může zvolit příklady z vlastního života. Obecné úvahy se ovšem týkají nás, opravdových lidí, nikoliv fikčních postav z našich ukázek.

logiku, snadno si však uvědomíme, že se obojí typy paměti musí vzájemně prolínat a podmiňovat. Epizodická paměť poskytuje nezbytný materiál k zobecňování pro účely sémantické paměti (nehledě na to, že každý úspěšný vklad do této paměti má za sebou konkrétní epizodu vkládání – každou svou vědomost jsem se musel jednou dovědět). A naopak, bez sémantické paměti bychom svým epizodickým příběhům stěží rozuměli, neuměli bychom je pojmově uchopit, zapamatovat si je a vyprávět.

Endel Tulving, průkopník studia epizodické paměti, ji charakterizoval jako

> evolučně mladý, pozdně vyvinutý a brzy ochabující paměťový systém orientovaný na minulost, pravděpodobně omezený na lidský druh. Umožňuje mentální cestování subjektivním časem z přítomnosti do minulosti, přičemž dovoluje znovuprožití vlastních dřívějších prožitků. [49]

Každý si denně na něco vzpomínáme, je to zcela běžná, až banální zkušenost. Je opravdu tak banální? Jak je možné, že se dovedeme přenést v čase kamsi nazpět a prožít si znovu něco dávno uplynulé, jako by se to stalo teď? Jiní vyšší živočichové si jistě dovedou leccos pamatovat, mohou se učit a cvičit, snad si i nějak uvědomují aktuálně prožívanou situaci, ale nezdá se mi, že by si mohli podle libosti vybavovat a znovuprožívat své dřívější příhody. Vlastně možná ano, údajně se jim zdají sny a komu se zdají sny, má asi též vzpomínky.

Celý běh lidského života se nám z odstupu může jevit jako rozsáhlé panoráma rozličných životních epizod (říkejme tomu tak: *panoráma*[14]). Jednu z těchto epizod prožívám v *přítomné*, čili *aktuální situaci*, ostatní si mohu jen lépe či hůře v rámci této aktuální situace zpřítomnit, vybavit z paměti (otázka empirické pravdivosti zde není relevantní). Toto zpřítomňování dříve prožitých epizod není vázáno nějakým pořádkem, záleží na mně, na co si právě teď chci vzpomenout, na co dřív a na co později. Některé vzpomínky (což je důležité) přicházejí samy od sebe – jako by mi je nějaké podvědomé či nevědomé pochody vsouvaly do zorného pole vědomí. Ale opět bez ohledu ke správnému (míním domněle správnému) časovému sledu.

Vybavujete si nějaký vlastní zážitek z minulosti (jako já ten z Kalifornie) a můžete si uvědomit, že vlastně prožíváte dvojí čas: čas v rámci epizody, *na níž* vzpomínáte, a aktuální čas v rámci jiné epizody, *v níž* na ni právě vzpomínáte. Vzpomínanou epizodu přitom prožíváte znovu, tentokrát jaksi virtuálně. A aby toho nebylo dost, je tu ještě třetí čas v pozadí, tušený čas celkového panoramatu vašeho života, v němž jakoby tato epizoda měla, spolu s ostatními epizodami, své předpokládané pevné místo a k němuž se též vztahují vaše obecné (nikoliv jen epizodické) zkušenosti, znalosti a zájmy.

Právě tato poslední úvaha mě přivádí k myšlence, že lze mluvit o dalším typu paměti, jakési *hybridní paměti*, v níž se epizodická paměť prolíná se sémantickou pamětí. Lze si myslet, že v ní jsou uloženy nasbírané a jaksi „zprůměrované“ osobní zkušenosti jednotlivce, které již ztratily vazbu na určitý epizodický prožitek, nicméně zachovávají si vztah ke svému nositeli. Vím (z *vlastní* zkušenosti), že v tramvajích se vyskytují revizoři (ale na žádného si konkrétně nevzpomínám), mám *své vlastní* kognitivní mapy („moje“ Praha na rozdíl od „vaší“ Prahy) a mám *své* dlouhodobější priority, touhy, obavy a specifické znalosti (s „nálepkou“, že jsou právě moje).

[14] Slovo si půjčuji z názvu knihy H.G.Adlera [1]. Viz též můj esej [18].

Zmínil jsem se, že mi jde o epizodické situace, které jsou vyjmuty z širších souvislostí. Co tím sleduji? Soudím, že některé zajímavé vlastnosti situací jako takových mohou lépe vyniknout právě tehdy, nemáme-li nutnost či nutkání je analyzovat v tom či onom kontextu, vztahovat je k tomu, co bylo dřív nebo co následovalo, co všechno mělo vliv na dění v jejich rámci atp. Můžeme se pak lépe zaměřit na prožitek epizodické situace sám o sobě, „uvnitř" jejího horizontu (viz sekci 4). Chceme-li (v první fázi) směrovat studium na konkrétní epizodické situace a jejich prožívání (později – ve druhé fázi studia – si můžeme dovolit zobecňování), je užitečné volit příklady, které jsou jakoby vytrženy z kontextu. Slovo „vytrženy" ovšem není vhodné, protože navozuje představu ostrých trhlin.

Vedle vzpomínek bychom neměli zapomenout i na situace ve snech. Prožíváme je jako samostatné epizody, chybí jim vzdálenější minulost a budoucnost a přímý vztah k bdělému životu[15], zůstává jen to nejnutnější ze sémantické paměti a jen velmi málo z hybridní paměti, navíc všelijak spletené a deformované. Přitom sám akt vzpomínání možný je:

Příklad 3. (Můj sen o botách[16])

> *Zjišťuji, že na nohou nemám boty. Hledám je, ale nikde nejsou, hledám i v přilehlých místnostech. Kde jsou má zavazadla? Měl jsem přece krosnu, nebo ne? Usilovně se snažím vzpomenout si, kdy a odkud jsem původně přišel a co jsem s sebou asi měl, ale ani přes velké úsilí si vůbec na nic nemohu vzpomenout.*

(S jiným příkladem snové epizody se ještě setkáme).

Jak se vzájemně liší vzpomínané, sněné a fikční epizodické situace? Zkusmo a v prvním přiblížení si lze všimnout řady rozdílů. O *vzpomínané* dávné[17] epizodě z vlastního života mohu vědět (1) to, co si z ní přímo a zřetelně pamatuji (v epizodické paměti), (2) co se mi podaří s větším úsilím dodatečně vzpomenout (stavím se k tomu, jako by to bylo někde v paměti ještě uschováno), (3) co vyplývá z mé obecné znalosti světa (sémantická paměť) a (4) co mi napoví „znalost" sebe sama (hybridní paměť). Se *sněnými* epizodami je tomu poněkud jinak, když je sním, vím jen to, (5) co snově přímo prožívám, (6) co si snově domýšlím, (7) co ve snu bezprostředně vyplývá z rezidua mé bdělé znalosti světa (fragmenty sémantické a hybridní paměti, které se protlačují i do snu) a také, (8) co mi tajemný autor snu pohotově nabídne v odezvu na moji (snovou) zvědavost. Konečně u *fikčních* situací mohu rozlišit tři vrstvy znalosti: vím (9) to, co je explicitně uvedeno v (daném úryvku) textu, (10) co vyplývá z širšího kontextu díla a ze znalosti autora (toto zde neberu v úvahu) či (11) z mé znalosti aktuálního světa (s tou však hodlám zacházet poněkud uvolněně), a konečně, (12) co si mohu (pokud chci) svobodně představit či domyslet ve fikčním světě (srov. též [8]).

Ve všech případech je důležité vždy umět přepínat mezi oběma zmíněnými tendencemi v reflexi, jak k 1PP prožívajícího subjektu (od čtenáře si to žádá schopnost vcítit se do hlavní postavy čili subjektu epizody), tak k 3PP (epizoda včetně jejího subjektu nahlížena z odstupu).

---

[15] Výjimkou jsou poměrně vzácné lucidní sny.

[16] Úryvek z delšího snu z noci na 29.1.2007.

[17] Slovo „dávná" má naznačit, že si již nepamatuji, co epizodě předcházelo, co po ní následovalo, a různé další souvislosti.

3 Modality situačního prožívání

V Příkladu 1 (Jezdec) si blíže všimněme popisované situace – nejdříve toho, co je v úryvku explicitně zmíněno. Především je tu čas (jízda *po celý den*) a prostor (jízda pustou *krajinou*; *nízké* mraky; ocitnutí se *před* domem). Prostor není prázdný, jsou v něm objekty, které se prostřednictvím sítě vzájemných pozičních vztahů podílejí na struktuře scény (*kůň* + *jezdec*; *mraky* + *obloha*; *jezdec* + *krajina*; *jezdec* + *dům*). A je tu ovšem i jakýsi děj, příběh či zápletka (*jízda*, která končí *ocitnutím se* před osudným domem). Tomu odpovídají první čtyři *modality situačního prožívání* (z perspektivy jezdce–vypravěče):[18]

(1) smysl pro *čas* a *trvání* (temporalita),

(2) smysl pro *prostor* a *rozlehlost*,

(3) smysl pro *strukturu scény* [*scene structure*] a

(4) smysl pro *zápletku* a *příběh* [*plot*].

Stejné modality lze určit i v příkl. 2 (Kalifornie): čas (*dlouho, nejméně hodinu*), prostor (*stojím nad*; *pode mnou*), scéna (*útesy*; *vlny*; *masivy skalního kotle*) a příběh (*vlny se opakovaně vrhají*; *nezdar*; *nový útok*). Podobně i v příkl. 3 (Můj sen o botách). Zdá se, že jsou to modality, které jsou do jisté míry univerzální (čili uplatňují se v mnoha, ne-li ve všech myslitelných situacích).

Můžeme ovšem pokračovat a identifikovat poněkud méně explicitní prožitkové modality, jako je smysl pro působení něčeho na něco (večerní slunce *vrhá* stíny / vlny se *vrhají* na skály; vždy po nezdaru se *zpění*), vědomí sama sebe jako prožívajícího subjektu (*já* na koni / *já* nad útesy / *já* bez bot), který si je vědom své tělesnosti (v odlišení od okolí) a navíc může být aktérem (agentem) dění – čili sám se cítit být původcem, autorem svých činů (*ovládám* koně / *stojím*; *napjatě sleduji* chování vln[19] / *hledám* boty). To vše je vzájemně prožitkově (fenomenálně) i logicky provázáno, nicméně je myslitelné tematizovat tyto další modality prožívání:

(5) smysl pro *působení* [*efficacy*] (jak se vzájemně ovlivňují některé prvky situace, zejména kauzální vztah příčina–následek),

(6) pocit *aktéra* [*sense of agency*] (moje vlastní působení, cítím se být původcem či autorem svých činů).

(7) pocit vlastní *tělesnosti* [*emodiment*] (včetně „vlastnictví“ těla),

(8) vědomí *sebe* (vlastního *já*, *jáskost*), jakožto účastníka situace [*self-awareness, selfhood*].

Zatím jsem neuvažoval situace, v nichž se vyskytuje druhá osoba anebo více osob; na tomto místě uvedu bez komentáře ještě další dvě modality (pojednám o nich v sekci 13 a 14):

(9) smysl pro *druhého* [*otherness*] (intersubjektivita),

---

[18] Zde i dále budu v hranatých závorkách uvádět anglické varianty některých termínů.

[19] I vnímání je druh aktivity.

(10) *socialita* (vztah subjektu ke skupině).

V uvedeném výčtu prožitkových modalit si ovšem nekladu nárok na úplnost, ba ani na „správnost“ – jde o zkusmou teoretickou konceptualizaci a pro různé účely se mohou ukázat jako vhodnější jiná rozlišení.[20]

Nikoliv náhodou jsem některé modality prožívání uvedl jako *smysl pro* něco (či *pocit* něčeho). Slovo „smysl“ (stejně jako „modalita“ – viz. pozn. 7) zde má ve velmi obecné rovině naznačit analogii se smysly jako modalitami percepce.

Čím více se v úvahách soustředíme na tu či onu modalitu, tím více se ukazuje, jak se všechny vzájemně prolínají, jak jsou provázány a jak na sobě závisí. Přesto si můžeme pojmovou analýzu usnadnit metodou, kterou nazývám „per partes“: v první fázi rozložit problém na několik komponent a zkoumat každou komponentu odděleně, vždy ovšem s přihlédnutím k existenci jiných komponent a jejich případných vlivů. Teprve v druhé fázi můžeme zaměřit pozornost na více komponent najednou a zjišťovat, jak vzájemně souvisejí a interagují. V našem přístupu hrají roli komponent jednotlivé modality prožívání a jejich vzájemná interakce vede k *multimodálnímu prožívání* (opět aluze na multimodální vnímání). Multimodální provázanosti se obecně nelze vyhnout, nebude však tématem této studie.

Z hlediska kognitivní vědy stojí za zmínku, že v některých případech je rozlišení modalit aspoň částečně podpořeno některými patologickými případy (jedno chybí, druhé je nedotčeno), fyziologickými znalostmi mozku, ontogenezí psychiky (různá stadia vývoje dítěte) a v neposlední řadě též běžnou introspekcí (která je v kognitivní vědě dnes již považována za legitimní zdroj poznatků). K poznání intermodálních souvislostí může přispět studium hraničních prožitků jednotlivých modalit.

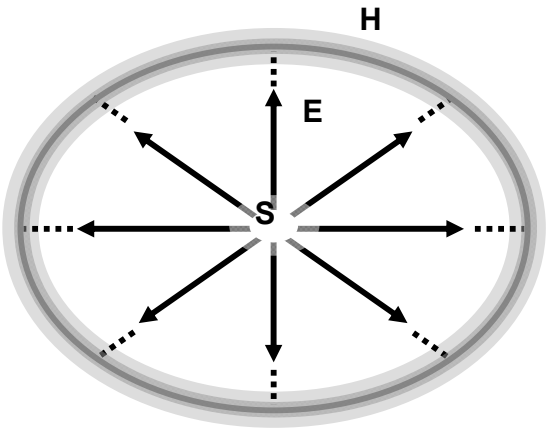
4 Pohled na horizont – hraniční prožívání a empirická věda

Z přirozené zkušenosti víme, že každá aktuálně prožívaná situace je nutně omezená už proto, že biologická konstituce našeho těla a našich smyslů nedovoluje překročit určité hranice poznatelnosti, některé pevné a jiné proměnlivé, ale vždy nějak přítomné. Víme, že prožívaný prostor je omezen do dálky i do hloubky smyslovou nerozlišitelností detailů; že čas přímo neprožíváme mimo přítomnou chvíli a i v ní nejsme schopni si uvědomovat příliš krátké a příliš rychlé děje; víme, že struktura scény závisí na naší předchůdné obeznámenosti s předměty, s jejich prostorovými vztahy a s tím, jak a k čemu se nám nabízejí (v gibsonovském smyslu, viz sekce 7), kam dosahuje akční rádius našeho fyzického působení a konečně že příběhy, jichž se účastníme, nikdy neznáme v plné rozmanitosti jejich dějových prvků – na otázku, zda a jak jsou tyto prvky implicitně propleteny vzájemným působením (srov. termín „zápletka“), můžeme zpravidla odpovídat jen na základě domněnek a dohadů.

V těchto všech dimenzích a v mnoha dalších jsou epizodické situace vždy nějak ohraničeny; pomyslnou hranici, kterou takto v některém či ve všech směrech tušíme,

[20] V práci [17] jsem vedle času, prostoru, strukturních vztahů, tělesnosti a sebevědomí uváděl (intencionální) jednání a smysl situace. Nyní považuji jednání za kombinaci modalit působení, pocitu aktéra, tělesnosti a vědomí sebe. Pocit smyslu (významu) situace nepovažuji za jednu z dalších modalit prožívání, nýbrž za něco, co je mimo či nad konkrétní situací, kterou přesahuje a vztahuje se k panoramatu života.

nazývám obecně *horizont* (obr. 1).[21] V této studii užívám pojem horizontu ve dvou stupních obecnosti, jednak pro dílčí horizonty vztažené k jednotlivým diskutovaným modalitám vnímání a prožívání (šipky na obr. 1) a jednak pro celkový, sjednocený *horizont epizodické situace* (rozostřená elipsa **H** na obr. 1).



Obrázek 1. Schematické znázornění horizontu epizodické situace
(**S** – pozice subjektu, **E** – oblast prožívání, **H** – horizont prožívání).

Pojem horizontu v sobě skrývá paradox: v jistém smyslu horizonty jsou všude a přitom neexistují, víme o nich a přitom je nevidíme. Kdykoliv si myslíme, že se nám nějakým trikem daří nějaký horizont překročit, vždy se vynoří jen o něco dál. „Lidská bytost neunikne horizontu," napsal filozof C. A. van Peursen, přičemž „horizont je lidské uchopení světa" ([51], s. 188, 195).

Dílčí horizonty mohou mít rozličný charakter. Speciálně pro prostorovou a časovou modalitu (později i pro některé další) lze rozlišovat mezi *distálním horizontem*, týkajícím se vnější periferie situace (dohled, celkové trvání), a *proximálním horizontem*, vztaženým k jemné struktuře vnímání času a prostoru, čili v jistém smyslu i „blízko" subjektu.

To, co jsem dosud napsal, je ovšem třeba číst opatrně. Příklady konkrétních horizontů a i obecný pojem horizontu jsem totiž formuloval jakoby z perspektivy třetí osoby (3PP, většinou skryté za oním zástupným „my") – jako o objektech, o nichž je možno mluvit jen z odstupu, z hlediska diváka (platí to i o schématu na obr. 1). Současně však horizont je pevně svázán s perspektivou první osoby, účastníka, subjektu situace (1PP). Jsou to dva komplementární diskursy, mezi nimiž je dobré umět pružně přepínat. Činíme to ostatně pořád, i v běžném životě, zvláště pak, když se rozhodneme reflektovat sebe sama v nějaké situaci. Je možné, že jedině díky takovémuto přepínání perspektiv jsme schopni si existenci horizontu vůbec připouštět a promýšlet.

[21] Pojem horizontu je metafora, zde je užita poněkud jinak, než ji užívají někteří fenomenologové. V mém pojetí má smysl užívat tři předložky: *před, na* a *za* (*před* horizontem, *na* horizontu, *za* horizontem).

Budeme se zde mimo jiné zajímat o *hraniční případy prožívání* [*liminal experience*]. Tím rozumím prožitky v „blízkosti“ horizontu té či oné modality prožívání.[22] Uvažujme například vizuální vnímání. Normálně, když něco vidím, pak *vím*, že to vidím – jde o nadprahové [*supraliminal*] vnímání. Má-li však vizuální stimul příliš krátké trvání (řádově několik málo desítek milisekund), nepronikne do vědomí, nicméně tělo (neuronová síť) stimul zachytí a může na něj i reagovat – je to případ podprahového „vnímání“ [*subliminal „perception“* [23]]. To jsou dosti známé věci; nás by však měla zajímat ona oblast *mezi* oběma případy (při délce stimulu kolem 100 ms), čili prahové či hraniční vnímání [*liminal perception*] – právě *na* proximálním horizontu vědomého prožitku. Důležité je toto: o vědomém prožitku mohu mluvit z perspektivy první osoby (1PP), přitom však nesmím (nemohu, neumím) překročit horizont (který je v našem vizuálním příkladu v jistém smyslu pevný, může ovšem záležet na jiných aspektech, únavě, náladě, kvalitě zraku apod.).[24] O všem dalším (o zachycení stimulu neuronovou sítí či o intervalech měřených v milisekundách) umíme mluvit už jen v diskursu třetí osoby (3PP – diváka, experimentátora, vědce).

Později (v sekci 11) se dostaneme k modalitě tělesnosti, pro niž americký filosof S. Gallagher [12] velmi obšírně a poučeně zkoumá kontrast mezi uvědoměnou představou (vlastního) těla – tělesným obrazem [*body image*] – a neuvědoměným tělesným schématem [*body schema*]. Píše:

> Fenomenologie se setkává s určitými přirozenými omezeními, když se ocitá před nefenomenálními procesy. Tělesné schéma v jistém smyslu představuje takové omezení.

Zkoumání zmíněného kontrastu Gallagher nazývá *negativní fenomenologií* ([12], s. 40).

Jak jsem již naznačil, budeme se zajímat o testování podobných kontrastů (anebo shod) mezi vědomými prožitky v 1PP a poznatky v 3PP; zdá se, že k takovým kontrastům (shodám) může docházet právě na horizontu prožívání té či oné situační modality.

Abychom se vyvarovali nedorozumění, rozlišme si aspoň konceptuálně dva typy hraničního prožívání (přesněji dva typy horizontu). Příkladem prvního typu je zmíněné prahové vnímání a rovněž prahové uvědomění si vlastního těla. Je to *vlastní horizont* [*intrinsic horizon*] – je přítomen v konkrétní situaci, je součástí *jejího* horizontu. Druhý typ hraničního prožívání (horizontu) lze označit jako *nevlastní* [*extrinsic*]. Existují případy, kdy pro některou modalitu rozlišujeme (rozumí se prožitkově, fenomenálně) různé stupně či intenzity prožívání, ale činíme tak spíše jen od situace k situaci. Lze si například představit rozsáhlou posloupnost situací, v nichž všech zvedám ruku, navenek stejně, avšak s klesajícím vnitřním pocitem mého chtění (ruka se může zvedat i mimovolně, bez mého chtění). V takové posloupnosti můžeme tušit nevlastní horizont chtěného pohybu. V opačném směru téže posloupnosti bychom zase mohli tušit jiný nevlastní horizont, horizont nechtěného pohybu.[25]

---

[22] Zde se nebudu zabývat patologickými kazuistikami, které by rovněž bylo možno považovat za specifické typy hraničního prožívání. Mnohdy mají i filosofickou relevanci mimo jiné v tom, že ukazují na vzájemnou provázanost či naopak nezávislost různých modalit prožívání.

[23] Uvozovky jsou proto, že o vnímání bychom správně měli mluvit jen v případě, že je uvědoměné.

[24] Zlepšíme-li své schopnosti (nasadím brýle), lze mluvit o „rozšíření“ horizontu anebo o „vtažení“ nových znalostí před horizont – podle toho, jak si představujeme metaforu horizontu.

[25] Takový případ střetnutí horizontů (a sám pojem horizontu) formalizoval P. Vopěnka [54].

Toto rozlišení typů horizontu je spíše orientační a závisí na tom, jak široce chápeme pojem situace. Při jiném pojetí mohou oba druhy splynout.

5 Smysl pro čas

Čas je nejobecnější podmínkou vědomí a tudíž (smysl pro) čas je nejdůležitější modalitou prožívání epizodických situací. Existuje rozsáhlá filosofická literatura zabývající se jak představami o povaze objektivního času, tak vnitřním vědomím času a nemám v úmyslu se zde zabývat různými pojetími času jako takového. Omezím se jen na některé aspekty časovosti, které souvisejí s naším studiem epizodických situací – speciálně nás bude zajímat vztah mezi dvěma pojetími času (1PP a 3PP) a rovněž hraniční prožitky času.

Konvenční, 3PP pojetí časovosti je založeno na představě absolutního a univerzálního objektivního času, který je homogenním, nekonečným a lineárně uspořádaným kontinuem. Vyjma kvantových a subkvantových teorií předpokládají všechny fyzikální vědy – a s nimi i ostatní přírodní vědy včetně (empirické) psychologie – že všechny přírodní děje a procesy probíhají v jednom společném, „světovém“ čase, reprezentovatelném osou reálných čísel (i s jejími vlastnostmi uspořádání, kontinuity apod.). Má to nejméně dvě výhody, jednak vlastnosti kontinua umožňují použít mocný aparát klasické matematické analýzy a jednak předpoklad jednotného času usnadňuje propojování teorií i empirických dat z rozličných domén vědeckého diskursu.

Naproti tomu pro filosofickou fenomenologii (počínaje Husserlem [24], srov. též [45]) a rovněž pro nově vznikající „vědu“ o vědomí (např. [10, 46, 50, 52]) je charakteristický důraz na subjektivní (1PP) zkušenost uplývajícího času. Východiskem není představa lineárního kontinua, nýbrž čistý prožitek aktuální, *uvědoměné přítomnosti*, zahrnující i právě odplývající (či odplynuvší) minulost a právě přicházející (či očekávanou) budoucnost (viz níže).

Tradiční pokusy sjednotit 1PP a 3PP na realistické bázi se setkaly s nezdarem. McThaggard [31] poukázal na neslučitelnost dvou modelů času, tak zvané série A (s 1PP triádou minulost–přítomnost–budoucnost) a série B (s 3PP relací uspořádání dříve–později), pokud se předpokládá realita času.

Kontrast mezi pohledy na čas v 1PP a v 3PP se promítá i do naivní reflexe časového charakteru epizodických situací. Subjektivně prožíváme čas jako proud událostí, které se vynořují z neexistující budoucnosti a skrze aktuální, okamžitou situaci odplývají do (opět) neexistující minulosti, dostupné už jen ve vzpomínkách. Ale současně jsme navyklí předpokládat existenci objektivního fyzikálního času s jehož pomocí lze (je myslitelné) prožité epizody chronologický pořádat a u každé měřit její trvání. „Přítomnost“ pak není než imaginární ukazatel pohybující se („rychlostí času“[26]) podél objektivní časové linie.

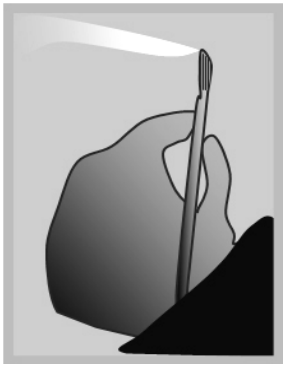
Lze tyto dva pohledy propojit, nikoliv jako dva formální modely času, nýbrž v reflexi naší zkušenosti času? Musíme přece připustit, že oba pohledy užíváme paralelně – přesněji řečeno, že vždy již nějak umíme přecházet od jednoho pohledu k druhému, *oscilovat* mezi nimi. Zkusme si obojí přechod ukázat na epizodě z Příkladu 2 (Kalifornie).

[26] Tento pojem užil (pokud vím jediný) Vopěnka [54a], s. 194, a má zřejmě smysl pouze při pokusu o propojení obou zmíněných pojetí času.

Vyjdeme-li z 3PP, celou epizodu (označme ji *E*) popisujeme objektivně, jako sekvenci podepizod (*útoků vln na masivy skalního kotle*), z nichž každá má nějaké trvání a následují jedna po druhé v rámci epizody *E*, kde jsou tedy lineárně uspořádány v segmentu objektivního času (trvajícím *nejméně hodinu*). Moje přítomnost, jako pozorovatele, jako by vůbec nehrála roli. Nyní však mohu uzávorkovat svou znalost objektivního času a uvědomovat si svůj prožitek (v 1PP) každé jednotlivé podepizody jako něčeho, co se odehrává v aktuálním „teď“. (V další úvaze ponechme stranou, že každá taková podepizoda má sama svůj vlastní strukturovaný vnitřní čas: *útok–nezdar–rozhoupávání*; toto uvažovat by neznamenalo než opakování téže úvahy o úroveň níž.) Co však je důležité: k prožitku každé podepizody náleží i matné, tratící se vědomí něčeho, co jí bezprostředně předcházelo (*rozhoupávání před útokem*), a stejně matné, probouzející se vědomí něčeho, co teprve přijde (*nezdar, zpěnění*). A tak se každá podepizoda v našem vědomí plynule přelévá do další podepizody a ta opět v další, z čehož se rodí subjektivní pocit plynutí času i v prožitku celé epizody *E*.

Opačný přechod, od 1PP k 3PP, spočívá v abstrakci z abstrakce [34]. Při první abstrakci vzniká z původního aktuálního prožitku jediného „teď“ představa šňůry takových „teď“ – „teď“ – „teď“, druhá abstrakce pak již vede k běžné 3PP představě fyzikálního času jako souvislé osy *t*.

Pokud bychom si chtěli zachovat aspoň částečnou subjektivitu času ve větším měřítku, museli bychom místo druhé abstrakce postupovat hierarchicky vzhůru, opakováním první abstrakce pro postupně složitější a rozpjatější epizody až po již výše zmíněné panoráma života, čili osobního životního času jednotlivce. Poznamenávám, že takováto hierarchická konstituce času nevyžaduje, aby epizody kterékoliv úrovně byly lineárně uspořádány (srov. [18]).



Obrázek 2. Malířova ruka (příklad atomické epizody)

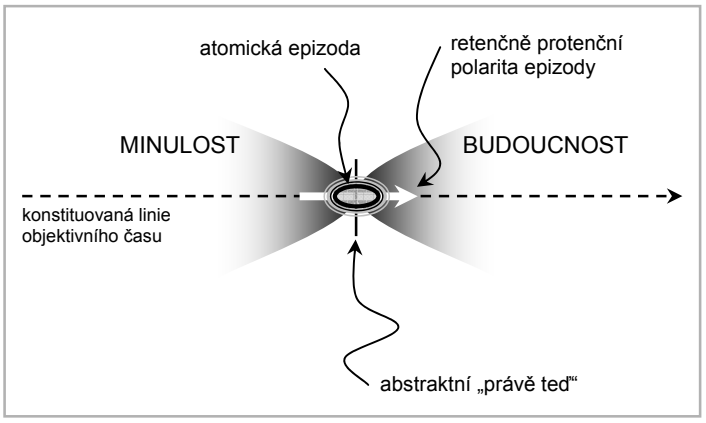
Nabízí se otázka hierarchické dekompozice směrem „dolů“, k menším měřítkům. Jak už víme, nelze pokračovat příliš dlouho, brzy bychom narazili na hranice lidské schopnosti vnímat krátké časové intervaly – čili ve smyslu předchozí sekce na proximální horizont prožívání času.[27] Lze se ptát na povahu časově „nejkratších“ prožívaných situací, které budu dále nazývat *atomické* epizodické situace.[28]

Na tomto místě je záhodno připomenout Husserlovu fenomenologickou teorii prožívání „vnitřního“ času [24]. Předreflexní prožitek „teď“ nelze považovat za nějaký časový bod (*okamžité* „teď“), ani za měřitelný interval (např. pomocí hodin). Jeho struktura sestává nejen z přímého bezčasového vněmu (*prezentace*), ale i z bezprostředního prožitku přidržení toho, co právě uplynulo (*retence*), a z očekávání toho, co právě přijde (*protence*); obojí s neurčitým horizontem. Představte si, že hledíte na reálnou (nikoliv zobrazenou) malířovu ruku se štětcem (obr. 2). Prezentace je abstraktní

[27] Je to příklad horizontu, který lze objektivními metodami měřit: podle typu prožitku je někde mezi 10 a 100 ms. Nepřímými metodami (zpomalený film) lze i tento horizont odsunout, při větším odsunutí se však ocitneme v cizím, nepřirozeném světě. Nebudu to zde uvažovat.

[28] V atomické epizodě proximální a distální horizonty smyslu pro čas splývají. Nezapomeňme ovšem, že sám pojem atomické epizody je spíše konceptuální abstrakcí než prožívaným fenoménem.

fáze „právě teď“ trojfázového intencionálního aktu vnímání situace, kterou současně prožíváme jednak ve světle doznívajícího uvědomění si *právě minulého*, již uskutečněného pohybu štětce zleva a jednak ve světle očekávání *nejblíže příštího* pohybu štětce směrem doprava.[29] Mluvíme o retenčně protenční struktuře atomické epizody, což spolu s pocitem kvalitativního rozdílu mezi oběma fázemi (odbytá minulost oproti nejisté budoucnosti) vytváří charakteristickou *polaritu* atomické epizody (obr. 3).



Obrázek 3. Retenčně protenční struktura prožívaného „teď“.

V našem přístupu lze tedy považovat atomické epizody za imaginární jednotky zkušenosti, jimž chybí (nadprahový) prožitek trvání, nikoliv však retenčně protenční polarita. Díky této polaritě a horizontové povaze retence a protence jsou atomické epizody připraveny k řetězení s jinými atomickými epizodami (což neznamená nutně přímý kontakt nebo překrytí) a tudíž i vytváření složitějších epizodických situací. Tuto úvahu lze ovšem opakovat i pro epizody s postupně rostoucí složitostí. Nahlíženo z 3PP (tedy z odstupu) si lze představovat, že při zmíněném „řetězení“ se polarity epizod nižšího řádu uspořádávají a vytvářejí vždy polaritu epizody vyššího řádu, což by mohla být i cesta, jak vykládat přirozené, vnitřní prožívání času jako něčeho, co uplývá, a považovat tento prožitek jako prvotní zdroj času konstituovaného v podobě souvislého a lineárně uspořádaného kontinua. Alternativně lze na pojmu atomické epizody založit i představu „zrnitého“ času [*granular time*] jako proudu atomických epizod na pozadí objektivního času ([19]).

Není divu, že smysl pro čas hraje mezi ostatními modalitami prožívání principiální a neodmyslitelnou úlohu. Čas je „univerzální podmínka bytí vůbec, ovšem čas ve smyslu tohoto původního vnitřního uplývání, nikoliv ve smyslu přírodního času partes extra partes.“ ([37], s. 103).

Onen konstituovaný lineární čas přírodních věd má ovšem řadu výhod, například umožňuje těmto vědám formulovat (v diskursu 3PP) přírodní zákony, měřit a popisovat dynamické děje a vytvářet srozumitelné kauzální domény. Nic také nebrání aplikovat

[29] Stejně trojfázové je i vědomí virtuálního pohybu, který si představujeme při *pouhém pohledu na obrázek* – to ovšem za předpokladu sémantické nebo hybridní znalosti pozadí: co je to ruka, štětec, barva, jaká je typická intence malíře apod.

tento čas – jakoby „zvenku“ – i na pozorovaný lidský subjekt. Pro naše další úvahy, v nichž se chci pokusit o analýzu modalit prožívání situací v „blízkosti“ horizontu, je užitečný i pojem *měřítka*, vycházející z 3PP.

Konkrétně si můžeme rozlišit následujících pět vrstev vztahu k času, a to podle zvětšujícího se měřítka (objektivního) času:

(1) *Fyzikální čas v malém měřítku* (řekněme mikrosekundy a méně – tedy včetně času subneuronálních procesů); čas v tomto měřítku zpravidla není uvažován při studiu vyšších kognitivních procesů. Každopádně není pro nás podstatné, zda fyzikální čas v tomto a menších měřítkách reprezentujeme jako ideální kontinuum či jako něco, co je v jemné struktuře je diskrétní.

(2) *Podprahový čas* (od mikrosekund k desítkám milisekund) – čas neuvědoměných fyziologických a neuronálních (subperzonálních) procesů.

(3) *Prahový čas* (zhruba desítky milisekund) blízko hranice (horizontu) vědomého prožívání. Fenomenálně jen retenčně protenční polarita atomických epizod.

(4) *Nadprahový čas* (od desítek milisekund po sekundy): zdánlivá přítomnost [*specious present*] u Jamese [25] (3 až 8 sec) či přítomný moment [*present moment*] u Sterna [46] (1 až 10 sec) – trvání nejkratších epizodických situací s narativním obsahem. V této vrstvě může docházet k emergenci pocitu časové kontinuity (plynutí).

(5) *Autobiografický čas* (hodiny, dny, roky,...) relevantní pro epizodickou a hybridní paměť (viz sekce 2).

Všimněme si, že zatímco vrstvy (1) a (2) jsou za proximálním horizontem fenomenální (1PP) zkušenosti času a jsou tudíž přístupny pouze vědeckému popisu (v 3PP),[30] ve vrstvách (4) a (5), pokud reflektujeme čas v 1PP – pak typicky jako *trvání*. Nejvíc by nás měla zajímat právě vrstva (3), kde se obě perspektivy mohou setkat, střetnout či splynout.

S určitou dávkou nadsázky se dá říci, že smysl pro čas (jako trvání) se rodí právě na horizontu uvědomění. Atomické epizody se postupně slévají (mimo jiné díky pocitu, že jsem to stále já, kdo je prožívá) v jeden proud a v nadprahové vrstvě tak může vznikat pocit (či chcete-li, iluze) kontinuálního toku časového vědomí, o němž jsem se zmínil výše. Již citovaný fenomenolog Shaun Gallagher píše:

> Zkušenost, místo toho, aby byla organizována v časovém proudu, spíše připomíná déšť dopadající na konečný povrch, kde se kapky zážitků sestřikují do kalužin smyslu, a ty jen občas vytvoří krátkodobý proud. ([10], s. 201).

Metafory „slévání“ a „proudu“ jsou úzce svázány s horizontovou povahou epizod (totiž s distálními horizonty v modalitě smyslu pro čas v rámci epizody). Fenomenálně (ale ani fyziologicky) neexistuje ostré zakončení jedné epizody, po němž by následoval ostrý začátek další epizody. Prvá může vplynout do druhé, může však mezi nimi být i mezera, ta by však nebyla přímo prožívána a byla by myslitelná jen jako abstrakce.

Při úvahách o času a časovém vědomí se s jakýmsi druhem „slévání“ setkáváme i na podprahové úrovni, kde je například dobře dokumentován sakadický pohyb očí, díky

[30] Měřítkové vymezení se u těchto dvou vrstev ovšem nevztahuje na domény diskursu odpovídajících vědních oborů (např. fyziky a neurověd).

němuž se krátké fixační periody (120 až 130 ms) střídají s periodami krátkodobé „slepoty“ během pohybu očí (s trváním do 100 ms); podprahové vjemy z fixačních period se nám ve vědomí slévají do jednoho celkového obrazu ([27], s. 63–64). Ale i na úrovni autobiografické dochází k něčemu analogickému: aniž bychom měli k dispozici kompletní vzpomínky na všechny epizody vlastního života, přesto získáváme pocit kontinuálního životního panoramatu [18].

6 Smysl pro prostor

Vedle smyslu pro čas je smysl pro prostor druhou nejvíce univerzální modalitou prožívání epizodických situací.[31] Mnohé z toho, co bylo v předchozí sekci řečeno o času, platí – *mutatis mutandis* – i o prostoru. Lze rozlišovat distální a proximální horizonty (v prvém jde o dohled a dosah, ve druhém o nerozlišitelnost). Při reflexi prožitku prostorovosti si rovněž bez obtíží představíme obě reflektivní tendence, jednu směřující k objektivizujícímu diskursu (v 3PP) a druhou zdůrazňující subjektivní prožitek prostoru (v 1PP) [23].

Existují též závažné rozdíly. Řeknu to obrazně: čas mne unáší, kdežto prostorem se potuluji sám, jsem v něm svobodnější a mohu se i vracet do míst, kde jsem již byl. Zajisté, pohyb a návrat nejsou myslitelné bez souhry času a prostoru, to však nic nemění na tom, že čas a prostor reflektujeme odděleně. Proto je považuji za dvě odlišné modality prožívání.

Při tendenci k 3PP se spíše opíráme o abstraktní geometrickou intuici trojrozměrného prostoru, v němž se epizodická situace odehrává, a v němž lze z obecných zákonitostí perspektivy objektivně určovat vzhled prostorových útvarů a jejich relativní, nicméně objektivní polohy vzhledem k subjektu. Uveďme si následující popis (fikční) situace:

Příklad 3. (Dvorky) [1a], s. 61.

> *Když jsem prošel průjezdem, octl jsem se na okraji prostoru, který měl půdorys pravidelného trojúhelníka; překvapil mě tím, že byl ještě rozsáhlejší, než jsem si představoval venku na ulici. Odvěsny trojúhelníka tvořily dvě fronty činžáků, přeponou byl vysoký násep železniční trati, zarostlý neproniknutelným tmavým křovím. Celý prostor byl rozdělený na dvorky patřící jednotlivým domům, ale ty byly od sebe odděleny jen nízkými zídkami, jež bylo snadné překročit.*

Vypravěč (v roli subjektu situace) se zde uchyluje přímo ke geometrickému slovníku (*půdorys pravidelného trojúhelníka*; *odvěsny trojúhelníka*; *přepona*). Jako by reálný prostor byl nahlížen očima geometra.

V geometrickém prostoru můžeme měřit, definovat transformace, kreslit mapy; každý objekt má své konkrétní místo, velikost a tvar. Navíc lze předpokládat, že geometrické vztahy jsou objektivní, univerzální a invariantní i při změně měřítka. Krátce řečeno, máme zpravidla v povědomí trojrozměrný euklidovský prostor. Geometrické pojetí prostoru umožňuje na jedné straně abstrahovat od egocentrické perspektivy diváka a

[31] V této studii se nezabývám čistě „kontemplativními situacemi“, v nichž je subjekt zcela pohroužen do svých myšlenek bez vztahu k okolnímu světu. V takových situacích nehraje prostorovost (na rozdíl od časovosti) svou obvyklou roli.

od horizontovosti, avšak na druhé straně umožňuje reprezentovat i tuto perspektivu (stačí vhodně zvolit souřadný systém a vhodnou transformaci), stejně jako částečně objektivní horizontovost (srov.: *octl jsem se na okraji prostoru*).

Jako situované lidské bytosti však neprožíváme své okolí primárně geometricky, aspoň pokud si to speciálně nenavykneme nebo se o to nesnažíme. Nastíním nyní možný scénář původního 1PP prožívání přirozené prostorovosti, který se však nevzpírá dodatečné interpretaci v diskursu 3PP [23].

Potlačím-li v dané situaci navyklou představu prostoru v 3PP, čili „v pohledu odnikud“, odhalí se mi prostor ve svém významu *přímo pro mne*, jako prostor *k něčemu*, k tomu, o co mi jde. Jako pro svobodnou bytost je to především *prostor k jednání* – k mé vlastní činnosti, k aktivnímu zasahování do situace.[32] Berme to zde v užším smyslu, jako jednání tělesné, říkejme raději fyzické (pohyb, zaujímání místa, prohledávání, setkávání s druhými, apod.). O prostoru pak říkáme, že je jednou „užší“, jindy „širší“, že je „něčí“ a že je kýmsi nebo čímsi někomu (mně) „poskytován“. Takto chápaný prostor úzce souvisí se svobodou volby – v každé konkrétní situaci se před subjektem otevírá rozvrh myslitelných alternativ, jeho vlastních možností. Budeme dále uvažovat speciální případ, kdy jednáním je *fyzický pohyb osoby* – subjektu prožívané situace – půjde nám tedy o *prostor k pohybu* ([23], kap. 2.; [15]).

Vypravěč v příkl. 4 je překvapen rozsáhlostí prostoru – již i to svědčí o subjektivním postoji, ježto jde o srovnání s očekáváním (dokud byl *venku na ulici*). Ale ještě „subjektivnější“ (a méně geometrické) je povšimnutí si překážek možnému fyzickému pohybu (*neproniknutelné křoví*; *nízké zídky, jež je snadné překročit*). Dvě věci tu stojí za povšimnutí: zaprvé, že jedna a táž situace je popisována v obojím prostorovém diskursu (3PP – geometrickém – a 1PP – k pohybu); zadruhé, že v tomto specifickém případu je zmínka o neproniknutelnosti a možnosti překročit více méně univerzální přes celý lidský druh, je tedy „objektivní“ v intersubjektivním smyslu (neplatí pro želvu nebo slona).

Ještě dřív, než jakkoli registruji vzhled nějaké scény, již nějak předvědomě rozumím základním možnostem svého pohybu – mohu se vydat vpřed nebo vzad, otočit se doleva či doprava, vystoupit výš, sestoupit níž. Vše vztahuji ke svému „zde“ – místu, v němž se právě nacházím (*octl jsem se na okraji prostoru*), a k mému „nyní“ – času, který prožívám jako přítomnost. Vědomí oněch možností je součástí prožívané přítomnosti (v obou smyslech slova „přítomnost“). Když se moje znalost situace rozšíří (uvědomím si *neproniknutelné křoví* a *nízké zídky*), horizont potencionální dostupnosti se roztáhne (jak hned uvidíme, tato představa je krokem ke geometrickému modelu prostoru k pohybu, čili k vyjádření v 3PP).

Vraťme se k oněm elementárním možnostem pohybu. K jejich uvědomění není rozhodující, zda se pohybovat opravdu mohu, chci, či musím. Co je však důležité, že o nich vím, že mám cit pro svou schopnost „dát se do pohybu“. To mi totiž poskytuje určitý prahový prožitek prostoru, který v určitých aspektech dokonce připomíná prahový prožitek času.

[32] *Jednáním* rozumím jakoukoliv aktivitu subjektu, přičemž (pokud si to nevyžaduje kontext) nepřihlížím ke stupni, v jakém jde o aktivitu uvědomenou. Častým případem je, že na vyšší úrovni subjekt jedná vědomě a intencionálně, přičemž na úrovni detailu má toto jednání nevědomé nebo neuvědoměné prvky. Blíže se o tom zmíním v sekci 10.

Zkuste si představit situaci, v níž jste zcela omezeni co do pohybu. Jste v mlze, ve tmě, pohřbeni v hlíně hluboko pod zemí. Vždy však něco máte: máte ono své „zde“, včetně jakéhosi elementárního, zárodečného ponětí o prostoru v bezprostředním okolí tohoto „zde“. Toto ponětí je založeno na vašem smyslu pro vlastní tělesnost a nutně k němu patří i smysl pro elementární *směry* či *polarity* myslitelného, byť (v dané situaci) nemožného pohybu: nahoru–dolů, dopředu–dozadu, doleva–doprava a ovšem i všechny jejich kombinace. Podobně jako u modality časového prožívání z předešlé sekce, pro níž okamžitá přítomnost (žité „teď“) neměla v našem pojetí žádnou extenzi, měla však polaritu, můžeme si i pro modalitu prostorového prožívání představovat proximální horizont jako jakési „zde“ bez extenze, jen s varietou myslitelných, byť třeba neuskutečnitelných možností „dát se do pohybu“. Takto chápané „zde“ budeme (ve shodě s [23]) nazývat *zónou elementární dostupnosti* – s důrazem na “elementární“, nikoliv na dostupnost. Protože v našem pojetí je neodmyslitelnou součástí každé epizodické situace její subjekt, má zóna elementární dostupnosti smysl v každé, i složité situaci.

V dalším kroku si můžeme představit svazek spojitých trajektorií potencionálního, avšak uskutečnitelného pohybu subjektu, které vyvěrají do všech směrů z jeho současné pozice, využívajíce možností, které scéna nabízí, vyhýbajíce se překážkám a respektujíce fyzické dovednosti subjektu. Tento svazek trajektorií nazveme *polem dostupnosti*. Toto pole je ohraničeno vlastním distálním horizontem, který je součástí celkového horizontu epizodické situace. (Tato úvaha je ve shodě s běžným užíváním slova „zde“ i ve vztahu k větším oblastem, ne-li k celému poli dostupnosti, zvlášť když toto pole má vymezené hranice uvnitř scény. Říkáme například „zde“ – před plotem, na rozdíl od „tam“ – za plotem.)

Dospěli jsme od subjektivního (1PP) pojetí prostoru k pohybu k pojmu pole dostupnosti a čtenář si jistě všiml že jsem opakovaně užíval typicky jazyka 3PP, dokonce geometrického (termíny jako „varieta“, „směr“, „trajektorie“, „pole“ apod.). Jak vidno, stěží se vyhneme geometrické intuici a geometrickým metaforám při snaze porozumět složitému světu. Současně jsme ovšem naznačili možný přechod od 1PP ke 3PP, aniž by se ztratily některé aspekty subjektivního hlediska (perspektivita, vztah k nabídkám prostředí). Při tomto přechodu byl důležitým východiskem prahový prožitek prostoru, což uvádím na podporu hypotézy, že hraniční prožívání nabízí most mezi pohledy v 1PP a 3PP.

Na tomto místě je vhodné, abych se zmínil o některých relevantních přístupech v psychologii a kognitivní vědě. Jako první lze připomenout klasickou gibsonovskou koncepci vizuálního vnímání, založenou na pojmu *ekologického prostoru* [13, 14]. Imaginární pozorovatel (či chcete-li, abstraktní referenční bod) se pohybuje tímto prostorem a jeho trajektorie je reprezentována souvislou sérií observačních bodů. Tato série je sdružena s třídou perspektiválních transformací v okolním optickém poli [*ambient optic array*], které poskytuje informaci o příležitostech k akcím, zvaných *nabídky prostředí* [*affordances*]. I když Gibson formuluje svou teorii výlučně v diskursu 3PP, jeho koncept ekologického prostoru je blízký našemu poli dostupnosti (přesněji třídě polí dostupnosti pro různé výchozí body).

Druhá relevantní teorie je založena na *enaktivním přístupu* v kognitivní vědě [53, 35]. Tato teorie klade důraz na vzájemnou souvislost vnímání a jednání, podle níž „naše schopnost vnímat nejen závisí, ale je i konstituována našimi senzomotorickými znalostmi.“ ([35], s. 2). Zkušenostní svět každého člověka (ale i vyšších zvířat) je svázán

s jeho *tělesným schématem* (viz sekce 11) a s jeho kategorizačními schopnostmi. Už i velikost těla hraje roli, protože vymezuje měřítka pro naši interakci se světem. Varela a spol. použili termín *vtělená kognice*, aby zdůraznili, že naše poznávací schopnosti jsou svázány s tím, že „máme tělo s rozličnými senzomotorickými schopnostmi, [které jsou] vloženy do širšího biologického, psychologického a kulturního kontextu.“ [53, s. 173].

Představuji si nyní sebe sama v roli subjektu situace v Příkl. 3 (Dvorky). Právě jsem prošel průjezdem a ocitám se v novém prostoru. Jsem překvapen jeho rozlehlostí (to však nechme stranou, stejně jako jakékoliv geometrické intuice) a představuji si, že se snažím porozumět scéně jako prostoru k pohybu. V pozadí mých myšlenek se rodí nové pole dostupnosti. Všímám si železničního náspu s neproniknutelným křovím, okolních činžáků a nízkých zídek, jež je snadné překročit. Jak si postupně uvědomuji všechny tyto objekty, moje pole dostupnosti se rozšiřuje, až nakonec zaplní celý dostupný prostor.

Vlastně nevnímám žádný prostor o sobě, vnímám jen *pohybově relevantní prvky*, jejich vzájemné pozice a vztahy. Rozumím jim jako takovým jen do té míry, v jaké o nich mám předběžné znalosti (srov. sémantická paměť) a jaké mám předběžné znalosti o svých vlastních pohybových schopnostech (srov. procedurální paměť). Řeknete-li: „Vida, prostor!“ je to ve stejné míře metaforou, jako když řeknete: „Vida, tady mohu leccos uvidět!“ To „vida“ totiž znamená daleko víc, než že něco je právě ve vašem zorném poli – máte totiž v povědomí i to, co byste (snad) mohli uvidět, kdybyste vykonali *pohyb* na jiné místo, ba co víc, jaké možnosti by vám nové místo nabídlo k dalšímu pohybu (obecně k dalšímu jednání). Opět se ukazuje úzký vztah mezi vnímáním a pohybem (jednáním).

Abyste cokoliv zahlédli nebo abyste se na cokoliv mohli podívat, musíte pro to něco *udělat* (přinejmenším pohnout očima či hlavou). V tomto smyslu je vizuální prostor rovněž prostorem *k* jednání, zde konkrétně *prostor k rozhlížení*. Obecně má percepce svou pohybovou komponentu a pohyb svou percepční komponentu, což je, jak jsem již poznamenal, hlavní myšlenka enaktivního přístupu k vnímání [35].

Z toho, co bylo řečeno, nevyplývá, že k pochopení prostoru jako pole dostupnosti je nutný reálný fyzický pohyb subjektu (vidomého, což je zde dobré předpokládat). Toto je důležitá poznámka, protože nám to umožňuje mluvit nejen o aktuálních epizodických situacích, nýbrž stejným způsobem i o situacích ze vzpomínek, snů, z literární a umělecké fikce a obecně o situacích, které se rodí jen v představách. Ve všech těchto případech lze reálný pohyb nahradit pohybem *virtuálním* ([23], sekce 6), čili pohybem *virtuálního subjektu* s *virtuálním tělem.*

Zkusme si to předvést na následující volné představě: abych porozuměl určité scéně, ať již reálné nebo imaginární, jako prostoru k pohybu, mohu si myslet, že mne zastupuje virtuální dvojník, který místo mne může pohodlně procházet rozličnými trajektoriemi možného pohybu. Virtuální dvojník nepodléhá fyzikálním omezením a může se pohybovat s libovolnou rychlostí a dokonce po mnoha trajektoriích souběžně. V duchu se vžívám do jeho pohybujícího se těla a tak poznávám prostor, který se mi v dané situaci otevírá – téměř jako bych tam opravdu byl. Výsledkem je, že získám celkové porozumění (či předporozumění) svým pohybovým možnostem v dotyčné situaci a – pokud jde o aktuální situaci – toto porozumění je mi pak k dispozici pro případ, že se rozhodnu opravdu se fyzicky přemístit.

Tímto způsobem snad lze uvažovat o schopnosti člověka (a možná některých zvířat) vytvořit si okamžité předvědomé porozumění potencialit vlastního pohybu i ve zcela nových a nečekaných situacích – porozumění, které musí předcházet jakékoliv

rozhodnutí se pohybovat. K tomu se hodí poznámka Merleau-Pontyho (v souvislosti s Wetheimerovým klasickým experimentem s převrácenou orientací obrazu scény):

> To, co hraje roli pro orientaci vnímané scény, není skutečná poloha mého těla jako předmětu v objektivním prostoru, nýbrž systém možných akcí, virtuální tělo s jeho fenomenálním „místem" definovaným jeho úlohou a situací. Moje tělo je kdekoliv tam, kde je třeba něco vykonat. ([32], s. 247–8)

Naše metafora virtuálního subjektu je sice zúžena na fyzický pohyb těla, nicméně nabízí obecnější interpretaci: místo „Moje tělo je kdekoliv, kde je třeba něco vykonat," můžeme říci, že virtuální tělo je kdekoliv, kde je *myslitelné* něco vykonat – a to dřív, než je jakákoliv úloha specifikována.

Zmínil jsem se o prahovém prožitku prostoru, založeném na pocitu „dát se do pohybu". Zóna elementární dostupnosti v dané situaci tak může odpovídat *vlastnímu proximálnímu* horizontu (viz sekce 4) prožívání prostoru k pohybu. K *nevlastnímu proximálnímu* horizontu by směřovala řada myslitelných situací s postupnou eliminací všeho, co by dovolovalo pohyb subjektu; jediné, co by zůstalo je zóna elementární dostupnosti.

Co lze říci o *distálním* horizontu prostorového prožitku? V kontrastu s předchozím případem by *nevlastní* distální horizont vyžadoval postupnou eliminaci všech překážek v pohybu. Limitní případ je těžké si představit subjektivně – absolutní uniformní širošírost, v níž by se (virtuální) subjekt všude pohyboval volně bez omezení (co by z toho měl?). Je to něco zcela nereálného, k čemu se můžeme zkušenostně blížit snad jedině v některých psychedelických stavech vědomí. V perspektivě 3PP je to ovšem blízké abstraktní představě vyprázdněného euklidovského prostoru – na nevlastním horizontu se tak 1PP a 3PP jakýmsi zvláštním způsobem setkávají.

Představa *vlastního* distálního horizontu prostoru k pohybu je naštěstí realističtější. Jde totiž právě o horizont pole dostupnosti pro danou situaci. Před tímto horizontem se uplatňuje 1PP, za ním již pouze 3PP; na horizontu obě perspektivy opět splývají. Obecně ovšem platí, že „lidská bytost neunikne horizontu"; vlastní distální horizont prostoru (a i jiných modalit) se projevuje jako takzvané *praktické* nekonečno [54] – nedosažitelná, neviditelná a nehmatatelná, ale vždy přítomná mez dohledu, dosahu, dostupnosti.

V této sekci jsem se soustředil na specifickou komponentu prostorové modality – prožívání prostoru jako prostoru k pohybu. Jen letmo jsem se zmínil o prostoru k jednání (pohyb vlastního těla je ovšem speciálním případem). Prožitek prostorovosti může záležet na mnoha dalších aspektech lidské situovanosti ve světě. Psychologové Cutting a Vishton ([5], s. 87) například rozlišují tři prožitkově různé prostory (v našem pojetí by to byly tři různé typy distálního horizontu). *Osobní* prostor [*personal space*], který sahá přibližně do délky natažené ruky, je prostorem intimním, kde mohu cokoliv snadno uchopit, a současně oblastí, při jejímž překročení ze strany druhé osoby se budu cítit ohrožen. *Akční* prostor [*action space*] je prostorem realizace vlastních záměrů, prostorem ústní komunikace a oblastí, kam mohu (řečeno obrazně i doslova) dohodit kamenem. Konečně prostor *dohledu* [*vista space*] rámuje celou scenérii, je relativně neměnným kontextem. (Autorům při tomto rozlišení šlo hlavně o klasifikaci různých zdrojů informace o trojrozměrném prostoru.)[33]

[33] Děkuji R. Šiklovi za upozornění na tuto práci.

Z hlediska filosofické fenomenologie je prožitkový prostor (v různých jeho variantách) vždy prvotní oproti prostoru geometrickému nebo fyzikálnímu. Cituji českého filosofa Jiřího Němce:

> Původní prostor je prosté „být u“ v jednotlivých svých způsobech (modalitách) – tj. tělesně (včetně smyslového kontaktu distančního – sluchem, zrakem), fantazijně, vzpomínkově, myšlenkově, touhou… a to celé hned nadvakrát (bděle, snově). Teprve na tomto základě (a nikdy bez něj) jsou možné různé abstraktní fyzikální prostory jako systémy souřadnic s různou strukturou (homogenní, nehomogenní apod.). ([34, s. 120)

Ve všech verzích prostorové modality je podstatné – a bylo to snad i vidět z výkladu v této sekci –, že tuto modalitu nelze zcela separovat od jiných modalit prožívání epizod, jako je smysl pro strukturu scény, pocit aktéra, vlastní tělesnost, ba i (jako v právě zmíněném případě) smysl pro druhého – „být někomu blízký“.

7 Smysl pro strukturu scény

Od smyslu pro prostor *per se* je možno odlišit poněkud jiný smysl, který se uplatňuje, když se v prostředí vyskytují nějaké předměty[34] (stěží si je představíme zcela prázdné) a když mezi těmito předměty jsou určité, hlavně poziční vztahy. Pak lze mluvit o *struktuře scény* – slovo *scéna* má zdůraznit právě tento aspekt (a ponechat prozatím stranou dynamické aspekty situace – těm se budu věnovat v následující sekci). Z kognitivních schopností je pro smysl pro strukturu reálné scény klíčové vnímání, a to především vnímáni vizuální.[35]

Všimněme si následující fikční scény z povídky A. Edwardsové:

Příklad 5. (Studovna)

> *Na odlehlém konci pokoje jsem skrz pootevřené dveře skříně zahlédl řadu geologických vzorků, anatomických preparátů, tyglíků, destilačních baněk a lahviček s chemikáliemi; na krbové římse vedle mne, mezi množstvím menších objektů, byl model sluneční soustavy, malý galvanický článek a mikroskop. Na všech židlích ležely věci a ve všech koutech pokoje byly hromady knih. Přímo na podlaze byly roztroušeny mapy, sádrové odlitky, listiny, nákresy a badatelské krámy všeho myslitelného druhu.* ([9], s. 126)[36]

Úryvek se zdá být téměř úplným popisem scény vyplněné předměty, popsanými v různých stupních typologické obecnosti (od *malého galvanického článku* až po *krámy všeho myslitelného druhu*), zpravidla s explicitním uvedením jejich pozice (*na krbové římse; na židlích; v koutech pokoje; na podlaze*). Popis můžeme považovat téměř za objektivní záznam (v 3PP), a to včetně odkazů k perspektivě vypravěče (*na odlehlém konci; vedle mne*). Úplnější popis v 3PP by ovšem mohl zahrnout i neviděné a neviditelné části scény.

[34] Či když v něm dochází k více či méně lokalizovatelným událostem, což zde nebudeme uvažovat.

[35] Nelze ovšem vyloučit ostatní senzorické modality, zejména hmat, avšak vizuálnímu vnímání a visuálním metaforám dávám přednost pro názornost a paradigmatičnost.

[36] Můj překlad.

Jak by se tedy lišil popis v 1PP? Uvedu aspoň pět kognitivně význačných podmínek vyvinutého smyslu pro strukturu scény, bez nichž by prožitek scény v 1PP byl stěží představitelný.[37] Jsou to:

(1) základní schopnost subjektu identifikovat vyskytující se předměty v jejich numerické identitě (má smysl je například počítat) a tudíž též schopnost rozlišit jeden předmět od druhého;

(2) znalost typologie předmětů (že krbová římsa je krbová římsa, že mikroskop je mikroskop apod.) – je jasné, že v tomto aspektu se uplatňují znalosti a zkušenosti transcendující aktuální situaci (sémantická paměť);

(3) schopnost identifikovat vzájemné prostorové vztahy předmětů (že něco je vedle něčeho, za něčím, uvnitř něčeho, na podlaze, co je blíž a co dál) – v tomto aspektu je důležité rozlišovat mezi vztahy, které jsou závislé na pozici prožívajícího subjektu, a vztahy, které jsou relativně „objektivní“ (i když i to jen v pohledu subjektu), – jde o rozdíl mezi *egocentrickou* a *allocentrickou* perspektivou;

(4) přístup k rozmanitým znalostem, které mohou vstoupit do vědomí subjektu (ať chtěně nebo samy od sebe) a jsou tak či onak relevantní pro danou scénu a předměty v ní (např. minulost věcí a jejich poukazy na nepřítomné věci a situace); sem by patřily i vlastní životní zkušenosti subjektu (hybridní paměť);

(5) v neposlední řadě pak porozumění významu prvků scény, ale i scény jako celku (např. co scéna pro subjekt znamená a co od něj očekává).

Zatímco v bodech (1) až (3) si lze poměrně snadno představit překlad do diskursu v 3PP, v bodě (4) a zejména (5) je to obtížnější. Dovedu si představit, že moje vlastní pocity by se ve studovně z Příkl. 5 (kdybych se v ní ocitl) podstatně lišily od hypotetických pocitů vypravěče (pro něho mohl mít hlavní význam třeba model sluneční soustavy, kdežto já bych se spíše zajímal o hromady knih). Všimněme si též, že v posledních dvou aspektech je struktura scény silně provázána s ostatními modalitami prožívání.

Již zmíněná enaktivní kognitivní věda zdůrazňuje souvislost mezi vnímáním a aktivitou subjektu (viz např. [53, 35]). V tomto pohledu obsah percepční zkušenosti je ve velké míře založen na senzomotorických znalostech [*sensorimotor knowledge*]. Určité části prostředí jsou pro subjekt nabídkami, přičemž takovéto “nabízení není čistě objektivním faktem, nýbrž je vymezeno relativně vůči systému, který je schopen určitých věcí v okolí užívat.“ ([12], s. 243). Řečeno jazykem enaktivismu, můj prožitek scény je cyklickým (recipročním) procesem, v němž moje senzomotorická znalost (soubor dovedností) průběžně interaguje s nabídkami scény. V tomto smyslu si scénu a její strukturu sami *zjednáváme* [*enact*].

Při vnímání scény je pro člověka charakteristická *orientovaná pozornost*. I některé z výše zmíněných podmínek relevantních pro kognici předpokládají, že subjekt přesouvá pozornost z místa na místo a z věci na věc (počínaje podprahovým sakadickým pohybem očí a intencionálním rozhlížením konče). To ovšem neznamená, že zbytek scény je zcela

[37] Omezuji se na snadno fragmentovatelné scény. Pro „souvislé“ scény (jako v příkl. 1) by bylo třeba následující body formulovat poněkud odlišně.

mimo pole vědomí. Zaměřím-li svou pozornosti na jednu věc, jsou ostatní věci *percepčně přítomny* ([35], s. 61), což znamená, že vím, že na ně mohu kdykoliv přesunout pozornost a tematizovat je. Jinými slovy, jsou v mém *marginálním* či *periferním vědomí* (v odlišení od tematizujícího vědomí; srov. [57], s. 62) – jsou *percepční komponentou fenomenálního pozadí* [6].

Běžný empirický výzkum v oblasti lidského vnímání – ale i některé filosofické úvahy – se často zajímají spíše o konkrétní vjemy konkrétních předmětů. Naproti tomu ekologický a enaktivní přístup je připraven zkoumat lidské prožívání scény jako celku a v jisté míře tak navazuje na gestaltismus 20. let minulého století. Tradiční otázky vztahu tvaru k pozadí a přesouvání orientované pozornosti z místa na místo se vracejí v nové podobě. Z novější psychologické literatury stojí v této souvislosti lze opět zmínit práci Cuttinga a Vishtona [5] o egocentrické vizuální kognici: struktura scény [*layout*] je vnímána prostorově jako výsledek integrace mnoha zdrojů informace.

V příkl. 5 vypravěč popisuje objekty v pokoji na první pohled v náhodném pořadí, záznam však můžeme číst též jako zprávu o jeho skutečném prožívání situace (možná se to od nás i čeká). Nejdříve si všiml skříně a jejího obsahu, pak krbové římsy, poté uviděl věci na židlích, knihy v koutech a nakonec vše na podlaze. Jeho pozornost se přesouvala – epizoda se stala jakýmsi mikropříběhem.

Prožitek scény jako celku obecně vyžaduje (až na triviální případy) určitou časovou extenzi, jejíž trvání si ani nemusíme uvědomovat. Atomická epizoda se proto omezuje jen na hraniční prožitek scény, zprostředkovaný lokalizovanou pozorností spolu s periferním uvědoměním si okolí. K periferním vědomí ovšem náleží i vědomí možnosti přesunout pozornost na něco jiného, čili možnosti si scénu intencionálně zjednat [35].

Vraťme se k našemu příkladu. Jak vidno, máme dvě možnosti. Buď čteme úryvek jako nezúčastněný divák „odnikud“, vytvářeje si tak objektivní (3PP) představu o scéně, můžeme si třeba kreslit i její obrázek. Anebo si, jako prve, představíme aktuální prožitek (v 1PP) osoby přítomné v popisované situaci. Ve druhém případě máme další volbu: buď se můžeme identifikovat se subjektem situace (v našem případě s vypravěčem) a přesovat svou imaginární pozornost po scéně přednostně ve vylíčeném pořadí a rozsahu, anebo se vtělit do myšleného virtuálního subjektu z předchozí sekce, který by si sice udržoval účastnickou perspektivalitu a horizontovost, měl by však úplnou svobodu se pohybovat libovolně po scéně a případně si domýšlet i ty detaily, které nejsou zmíněny v původním popisu. Tato poslední možnost vlastně tvoří opět intuitivní most mezi 1PP a 3PP.

Připomeňme si příkl. 4 z předchozí sekce (Dvorky). Čtete vyprávění a scéna se vám postupně otevírá, můžete si představit svého virtuálního dvojníka, ocitnuvšího se na konkrétním místě (*za průjezdem, na okraji prostoru ve tvaru trojúhelníka*). Z tohoto místa byste však stěží poznali strukturu celé scény (*dvě fronty činžáků; násep železniční trati; dvorky oddělené nízkými zídkami*), pokud byste neměnili směr svého pohledu, ba i své stanoviště v rámci trojúhelníkového prostoru (zkuste si scénu nakreslit).

V enaktivním přístupu má přítomnost subjektu na scéně závažnější úlohu, než že by jen určovala výchozí bod perspektivního vnímání. Je totiž rovněž výchozím bodem zasahování do děje – a hned se objevuje vazba k další modalitě, ke smyslu pro zápletku či příběh.

8 Smysl pro zápletku a příběh

V našem úvodním příkladu (Jezdec) se na první pohled nic dramatického neděje, jezdec na koni (a současně fiktivní vypravěč) jede celý den (*neobyčejně pustou krajinou*), dokud se neocitne na specifickém místě (*před melancholickým domem Usherů*). Ale právě tento závěr zřejmě otevírá nějakou zápletku (aspoň pro toho, kdo zná Poeovy povídky). Vím, že příběh je psán pro nás, čtenáře, kteří jsme tu jen jako diváci sledující jezdce (v 3PP – „jeho" jízdu). I jako divák–čtenář se však mohu empaticky vžívat do pohledu a možných pocitů jezdce, hlavní to osoby této epizody (líčené ostatně v první osobě singuláru). Přepišme si (s jistou dávkou fantazie) tuto epizodu v 1PP:

Příklad 6. (Jezdec – varianta**)**

> *Po celou cestu jsem pohroužen do svých myšlenek, nenechávaje se rozptylovat pustou krajinou a tísnivými mraky nade mnou – až, hle, s večerními stíny před sebou vidím melancholický dům Usherů. Co bude dál? Jak se věci vyvinou pro mne, co se mi odhalí, jak já se zachovám?*

Takto se bezpochyby do situace *zaplétám*, situace mě plně absorbuje. Slovo „zápletka" pak lze chápat ve velmi obecném smyslu, nejen v běžném smyslu dramatické zápletky. Vlastně ani není třeba v našem kontextu pojmově rozlišovat zápletku a příběh.

Úmyslně jsem zvolil příklad, který je zprvu (zdánlivě) chudý na děj. Jistě si dovedete představit, že jste do nějaké situace zataženi daleko více, kdy je i na vás, jak se zachováte a zda tím něčeho dosáhnete (anebo něco pokazíte). Například do situace z příkl. 3 (Můj sen o botách) byste se na mém místě mohli cítit velmi zataženi, pokud by vám (tolik jako mně) šlo o nalezení bot. Stupeň vaší angažovanosti může být větší či menší – větší nebo menší může být i stupeň vašeho *uvědomění* si vaší angažovanosti. (Všimněte si, že teď píši v druhé osobě, i když pro mne je můj čtenář třetí osobou, vy pro sebe zase první osobou, aspoň pokud jste si nějaký konkrétní zážitek vybavili. Vyprávění svého a cizího příběhu nemusí být navenek příliš odlišné, což nijak nesnižuje závažnost rozdílu.)

Pokud jde o vztah mezi 1PP a 3PP, je v prvním přiblížení u modality smyslu pro zápletku (příběh) podobný témuž vztahu u modality smyslu pro scénu, aspoň pokud se pohybujeme na „povrchu" příběhu. Nerozhoduje, zda je situace sama o sobě dynamická: tak v příkl. 2 (Kalifornie) nelze o dynamičnosti pochybovat, nicméně já sám, jako subjekt situace, v ní příliš angažován nejsem. Věc se ovšem stává složitější v případě, kdy je subjekt do příběhu zatažen podstatně víc. Pak se totiž dostává do hry i jeho bezprostřední odpovědnost za to, jak do děje zasáhne.

Co by mohlo být hraničním (prahovým) prožitkem v modalitě smyslu pro zápletku – čili blízko horizontu epizody? Pokud bychom se zajímali o *nevlastní* horizont, stačilo by si představit limitní situaci po eliminaci jakékoliv zápletky, situaci bez jakéhokoliv příběhu, který by stál za řeč. Mohla by to být například nějaká statická scéna, kde se nic neděje, kde není žádný rozdíl mezi divákem a účastníkem. V příkl. 4 (Dvorky) a 5 (Studovna) jsme mohli registrovat různá výskytová jsoucna, typově je určovat, zjišťovat jejich vzájemné polohy a případně si je představovat z rozmanitých perspektiv pohledu,

ale o příběhu jako by nebyla řeč. Ale je to vůbec možné? Je nesporné, že si vždy nějaký příběh můžeme dosadit.[38]

Daleko zajímavější by mohly být případy, kdy by se dalo hovořit o *vlastním* horizontu, čili tematizovat nějaké rozlišování *uvnitř* situace, které by svým gradientem poukazovalo k pomyslnému horizontu. Takový horizont by se zřejmě netýkal složitosti zápletky jako takové, nýbrž stupně, v jakém je subjekt epizody do zápletky zapleten, jak je v ní angažován. Dle naší hypotézy by hraniční případ angažovanosti měl poskytnout půdu pro splynutí perspektiv v první osobě (1PP) a ve třetí osobě (3PP).

Předveďme si to na úryvku z románu současného anglického spisovatele Iana McEwana „Pokání" (hlavní osobou úryvku i celého románu je třináctiletá dívka jménem Briony):

Příklad 7. (Briony u okna)

> *Otočená k oknu se* [Briony] *zadívala na poklidné řadové domky, do kterých se opíralo slunce, a na cestu, kterou přišla od High Street. Překvapilo ji, že se jí ještě nechce odejít. Přestože* […se] *bála, co dalšího ještě přijde. Pozorovala stařenu v tlustém zimníku, i když bylo takové teplo. Šla po druhém chodníku a na vodítku vedla nemocně vyhlížejícího břichatého jezevčíka. Cecilia a Robbie si teď tiše povídali a Briony se rozhodla, že bude jejich soukromí respektovat a neotočí se od okna, dokud na ni nepromluví. Uklidňovalo ji pozorovat tu ženu, jak otevírá vrátka, s nervózní precizností je za sebou pečlivě zavírá, potom se na půl cesty s obtížemi shýbá, aby vytrhla nějaký plevel z úzkého záhonu, který se táhl po celé délce cestičky. Pes k ní přitom připajdal a olízl jí zápěstí. Žena se psem vešla dovnitř a ulice se znovu vylidnila.* ([30] s. 298–9)

Pro tuto situaci je klíčovou rekvizitou právě okno, které je předělem mezi tím, co je daleko, a tím, co je blízko. Dění *za oknem* je daleko, nevyžaduje od Briony žádnou činnost. To již neplatí o dění *před oknem*, které se jí bezprostředně týká (*bála se, co dalšího ještě přijde*) a do něhož navíc může, rozhodne-li se, aktivně zasáhnout. Přesto scénu za oknem sleduje dlouho (tak dlouho, že stařena mezitím *stihla projít cestu od druhého chodníku skrze vrátka, zastavit se u plevele a vejít do domu*) a všímá si přitom podrobností (*tloušťky stařenina zimníku, nemocného vzhledu jezevčíka a jeho břichatosti, nervózní preciznosti stařeny při zavírání vrátek, tvaru a umístění záhonu, psího olíznutí*). Jako by se v pohledu do dálky Briony cítila osvobozena od delikátní situace před oknem.

Popsaná situace tedy sestává ze dvou scén oddělených oknem. O (dílčím) příběhu ve scéně za oknem se čtenář, – stejně jako Briony – dovídá celkem dost, zato o (dílčím) příběhu před oknem ví jen jedno: že neví – stejně jako Briony –, co vlastně si Cecilia a Robbie spolu říkají (situační znalosti čtenáře se zde příliš neliší od Brioniných). Nicméně cítíme, že situace před oknem je pro Briony naléhavější, dokonce ji činí *bezprostředně odpovědnou*. I nečinnost může zasáhnout do příběhu.

[38] Teoreticky si lze dokonce domýšlet nejméně čtyři logicky odlišné typy příběhu, což lze ukázat třeba na našem příkl. 4 (Studovna). První jsem už prve popsal: mikropříběh subjektu–vypravěče, seznamujícího se s dotyčnou situací; druhý typ příběhu transcenduje danou fikční situaci: kdo, kdy a proč si takto zaplnil pokoj různými věcmi; třetí je metapříběh empirického autora: kdo, kdy a proč to všechno vymyslel a sepsal; konečně čtvrtý příběh je prožitek čtenáře při jeho setkání s textem. Nemyslím však, že by takovýto extrémní nadhled nějak přispěl k našemu tématu.

Dvě výhrady na okraj: za prvé, rozlišení vzdáleného dění (nevyžadujícího žádnou činnost) od blízkého nelze zobecnit na všechny situace – i za oknem se může přihodit něco důležitého. Za druhé, „bezprostřední“ odpovědností rozumím spíše stav mysli souběžný s jednáním a tudíž náležící k dané epizodě. Nepopírám, že nepřímá či časově vzdálenější odpovědnost hraje v lidském životě podstatnou roli, uvažovat ji by však překročilo téma této studie.

Obrázek 4. Katsušika Hokusai: Fudži za bambusovým hájem.

Čtěme znovu ve snaze vžívat se do Brioniny perspektivy první osoby. Nezdá se vám (tak jako mně), že vy sami stojíte u okna a spolu s Briony – takřka jejíma očima – sledujete stařenu, jezevčíka a vše, co se děje venku? Nic statického, i venku se odehrává příběh, jeho protagonistkou však není Briony, nýbrž stařena (pomineme-li jezevčíka). Takže vaše perspektiva (jakožto diváka – v 3PP) se neliší od Brioniny perspektivy (jakožto účastníka – 1PP). Zcela v souladu s naším očekáváním splynutí obou perspektiv.

Naproti tomu scéna před oknem, kde je situace (pro Briony) daleko složitější, je příliš blízká a tudíž (navzdory perspektivě) nedovoluje takový odstup. Ostatně zde (před oknem) vůbec nemáte co dělat – nemůžete a nesmíte se plést do děje. A navíc zápletka, k níž zde (před oknem) dochází, je příliš propletena se zcela jinou situací, tentokrát *uvnitř* Brioniny duše (řečeno s dávkou metaforické licence). Což jste schopni být v kůži Briony vstříc delikátní atmosféře *před* oknem?

Ke struktuře scény tedy náleží i další dimenze, dimenze *dálky*. Gallagher například píše:

> Vnímaný svět získává obrysy skrze možnosti mého vnímajícího těla, těla, které je lidské a je schopno poměrně daleko dohlédnout, může se však dotknout nebo uchopit něco, co je poměrně blízko. ([12], s. 140)

Význam rozdílu mezi dálkou a blízkostí v epizodické situaci si ukážeme na jednom příkladu. Na obr. 4 není na první pohled nic nezvyklého. Skrze bambusový les prosvítá obrys vzdálené hory Fudži. Nevím, jak vám, ale mně tento obrázek vždy přiměje k

jakémusi elementárnímu prožitku distance mezi blízkým a vzdáleným, zdůrazněné neviditelnou, jen tušenou bezednou propastí mezi skupinou bambusových kmenů v popředí a horou v pozadí.[39]

Do epizodické situace uvedené na začátku naší studie (příkl. 1 – Jezdec) se polarita mezi blízkým a dalekým promítá do perspektivy jezdce (subjektu situace) poněkud jinak: blízkým je mu jeho kůň a zvolna se měnící výsek krajiny, kterým právě projíždí (a nakonec i dům Usherů), dalekým je celá ostatní krajina, jednak ta, kterou již projel, a jednak ta, která na něj teprve čeká.

Blízké a daleké – to jsou jen dva póly vzájemného vztahu: jedno nemůže být bez druhého. Totéž platí o dvou modalitách, pro prostor a pro strukturu scény.

9 Smysl pro působení

Ve velmi zjednodušujícím pojetí lze v každém příběhu vysledovat určité elementární události, kdy něco *způsobuje* či *ovlivňuje* něco jiného – aspoň kdy *nám* se to tak jeví. Nejčastěji pak prvé pokládáme za *příčinu* druhého a druhé za *následek* prvého, přičemž jejich vzájemný (typicky asymetrický) vztah považujeme za případ vztahu *kauzálního*. O něm zpravidla předpokládáme, že je podložen nějakou obecnou (ať už známou či neznámou) pravidelností či zákonitostí – obecně řečeno jsou kauzální vztahy hlavní, ne-li jediné vztahy, na nichž jsou založena veškerá vědecká (a tedy 3PP) vysvětlení.

Protože nám zde jde i o subjektivní (1PP) prožívání, mluvím raději o obecnějším, a proto méně specifikovaném vztahu *působení* [*efficacy*], jehož je kauzalita v běžném smyslu jen speciálním případem. Chci totiž zahrnout i případy, kdy v dotyčném jevu tušíme vůli či svévoli nějakého (přítomného nebo nepřítomného) aktéra, nebo naopak jakési spontánní tíhnutí či směřování něčeho k něčemu (srov. pojem propenzity u Poppera [40], Sec. 27), nebo konečně, kdy dotyčný jev prostě očekáváme, aniž bychom jej měli podložen nějakou, třeba naivní teorií.

Proč tato obecnost? Jde mi totiž o přirozený, nepoučený a nikoliv nutně reflektovaný pocit, že některé prvky se v dané situaci tak či onak ovlivňují, zatímco jiné jako by k sobě byly netečné. Tento vágní pocit nás posléze vede k rozlišování mezi ději, které připouštějí možnost vysvětlení (což ještě neznamená, že je umíme nebo chceme vysvětlit), a ději „čistě“ nahodilými.[40] Či chcete-li, mezi jevy, kterým rozumíme, a proto se jim nedivíme, a jevy, které nás mohou překvapit (což je ovšem poněkud jiné rozlišení).

Začnu opět fikčním příkladem. Jistě vám bude povědomý.

Příklad 8. (Alenka)

> *Alenka si s úžasem všimla, že jakmile oblázky dopadnou na podlahu, hned jsou z nich koláčky, i dostala šťastný nápad. “Když sním jeden koláček, jistě to zapůsobí na mou velikost, a protože větší už být nemohu, budu nejspíš menší.* ([4], s. 28-29)

[39] Scénu na obr. 4 lze považovat též za epizodickou situaci. Stačí si představit sebe v roli virtuálního subjektu (není to obtížné).

[40] Rozlišování velmi relativní, závislé na doméně diskursu a na našich znalostech světa.

Alenka zpozorovala jistý jev (*změnu oblázků v koláčky*)[41] a ač to nepovažovala – soudě z jejího úžasu – za něco zcela běžného, musela ji okamžitost změny (*jakmile...hned*) a multiplicita (*oblázků* bylo více) napovědět, že tu může jít o vztah působení, možná i příčinného (*dopadnutí oblázků* jako příčina, *změna v koláčky* jako následek). Tento pocit ji pak přivádí k domněnce, že jiný jev (*snězení koláčku*) může také nějak zapůsobit (*na její velikost*), dokonce to (z nejasných důvodů) považovala za jistotu (*jistě to zapůsobí*). Na základě této epizodické zkušenosti by stěží dospěla k nějaké obecné teorii, o to jí však nešlo, potřebovala se jen zmenšit.[42] Obdobně si můžeme všimnout případů zjevného i méně zjevného působení v některých našich dřívějších příkladech (*nezdařený výpad vln* působí *zpěnění*; *pozorování ženy na chodníku* vede k *uklidnění Briony*; *žena vešla dovnitř*, čímž *se ulice vylidnila)*.

Zajímáme se o hraniční případy (na prahu uvědomění), a proto stojí za zmínku Michottovy psychologické experimenty již ze 40. let [33]. V jejich pozdější verzi dle Shallice [44] (viz obr. 5) testovaná osoba sleduje pohyb dvou disků na obrazovce. První disk se pohybuje rovnoměrně dokud se nedotkne druhého (stojícího) disku, načež se zastaví; po určité prodlevě se druhý disk dá do pohybu a pokračuje ve stejném směru jako první disk (Obr. 5a). Ukázalo se, že při trvání prodlevy pod 70 ms má pokusná osoba pocit přímé fyzické kauzality, zatímco při prodlevě mezi 70 a 140 ms má pocit skrytého nepřímého působení a nad 140 ms dokonce žádnou souvislost nepociťuje (Obr. 5b). Jde tedy o konkrétní případ, kdy byl (v 3PP diskursu objektivního měření subjektivního pocitu – tedy 1PP) identifikován horizont modality (subjektivního) smyslu pro působení.[43] V našem pojetí (sekce 4) je to horizont nevlastní, protože různá trvání prodlevy odpovídají různým situacím.

Smysl pro působení v hraniční podobě prozrazuje úzkou souvislost dvou modalit, totiž smyslu pro čas (proximální horizont) a smyslu pro působení (dokonce v podobném měřítku času). Naproti tomu se nezdá, že by tato modalita přinesla něco podstatně nového pro dualitu perspektiv v porovnání se složitějším smyslem pro zápletku. Pokud se ovšem nezeptáme, zda se v roli působitele (příčiny) neocitá právě prožívající subjekt sám. To by ovšem mělo zásadní význam a proto volím tento případ jako samostatnou modalitu prožívání.

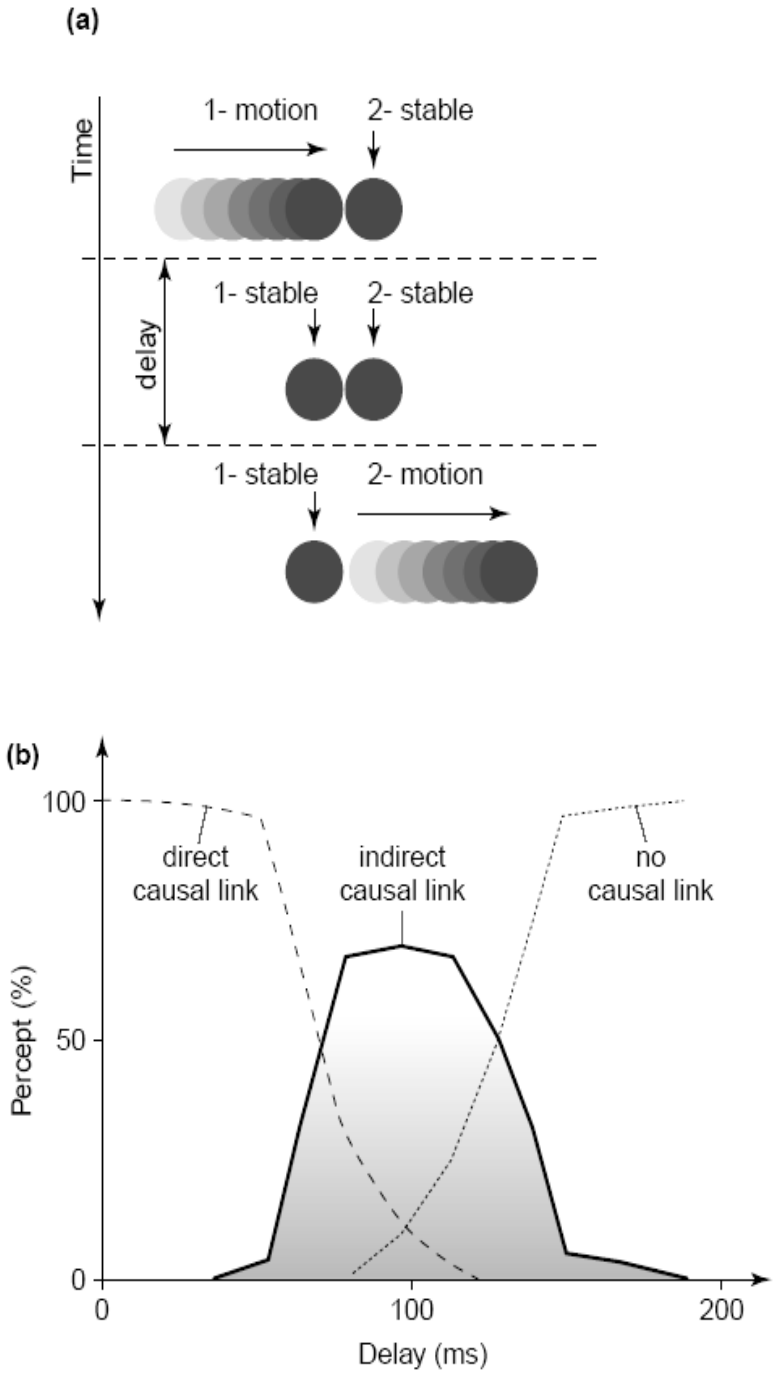
10 Pocit aktéra

Jestliže si všimnu, že pád oblázku na podlahu způsobuje změnu oblázku v koláček, a to bez jakéhokoliv mého přičinění, pak to budu prožívat za něco jaksi jiného, než když se *já sám* svobodně a cílevědomě rozhodnu koláček sníst, *abych* si tím způsobil změnu velikosti. Pomineme-li méně zajímavé případy nechtěného působení (např. když zatěžuji židli pod sebou nebo když překážím ve dveřích), jde tu o velmi závažnou věc: o *pocit aktéra* [*sense of agency*] neboli *autorství akcí* [*sense of authorship*] – že jsem to *já*, kdo je

[41] Otázka, zda jevy a obecně situace uváděné v příkladech jsou v reálném světě možné, není pro naše úvahy naštěstí relevantní.

[42] K čemuž použila půvabnou logiku: působí to na mou velikost + nemohu být větší, ergo budu menší.

[43] Stojí za zmínku, že vědomí toho, že jde o virtuální a nikoliv fyzické objekty, nemělo vliv na výsledky. Podobné experimenty byly později prováděny v různých obměnách, například změna směru pohybu či změna barvy druhého objektu.



Obrázek 5. Pocit kauzálního působení: (a) časový režim, (b) typ prožitku v závislosti na prodlevě. (dle Shallice [44]).

intencionálním původcem nějaké události, kdo něco (vědomě a ze své vůle) způsobil, kdo zasáhl do děje a kdo pociťuje odpovědnost za tento zásah.[44] V předchozím příkladu dostala Alenka nápad a podle něho se rozhodla jednat – cílevědomě zasáhnout do situace (*požitím koláčku ovlivnit svou velikost*). Což i vzápětí realizuje:

Příklad 9. (Alenka – pokr.**)**

> *Snědla tedy jeden koláček a nemálo se zaradovala nad tím, jak se valem scvrká.* ([4], s. 29)

Jak vidno, po mentálním aktu (*nápadu*) následuje fyzický akt (*požití koláčku*), který má navíc, jako každý intencionální akt, tři fáze, chtění, provedení a uvědomění si výsledku. Mezi nápadem a fyzickým aktem je jistá závislost, na niž vypravěč poukazuje jinak nenápadným slůvkem „*tedy*"; v našem pojetí lze tuto závislost považovat za případ obecného vztahu působení (mentální působení), zde nám však jde hlavně o doprovodný prožitek aktéra. Pak je tu ještě uvědomění si výsledku, hlavně zda byl úspěšný (*zaradovala se nad tím*), což je další případ vztahu působení.

Ostatně již v nápadu jako takovém je (v našem příkladu) obsaženo porozumění sobě samému jako *možnému* aktérovi (aktérce). Podívejme se však, jak je pocit aktéra svázán s fyzickým aktem samotným, především pak s pohyby těla (jako třeba zvedání ruky).

Existují psychologické testy na určení hranice mezi pocitem aktéra a pocitem, že dění není subjektem nijak ovlivněno. Testovaný subjekt sleduje svůj pohyb na počítačovém obraze, který je jednou více, jindy méně zpožděn. Je-li toto zpoždění pod 150 ms, subjekt považuje sledovaný pohyb za svůj vlastní, jinak jej za vlastní nepovažuje. Ukazuje se tak jeden specifický horizont pocitu aktéra, a to vzhledem k času; podobně jako při prožitku kauzality, jde o horizont *nevlastní*.

Pokud se naproti tomu zajímáme o *vlastní* horizont pocitu aktéra, je situace jiná. Chtěli bychom zkoumat, zjednodušeně řečeno, „kde" je pro konkrétní uskutečněnou akci hranice mezi její intencionální komponentou na jedné straně a mimovědomou, neosobní [*subpersonal*] tělesnou motorikou na straně druhé.

Opět nabízím ukázku fikční situace z již dříve citovaného románu I. McEwana.[45] Třináctiletá Briony právě sedí na podlaze dětského pokoje a prohlíží si ruce.

Příklad 10. (Brionin prst)

> *Zvedla jednu ruku, sevřela prsty a podivovala se, tak jako dříve už vícekrát, jak k této bytosti, tomuto stroji na uchopování, masitému pavoukovi na konci své paže, která ji na slovo poslouchá, přišla. Nebo žije nějakým svým vlastním malým životem? Ohnula prst a narovnala jej. Mystérium tkvělo v okamžiku, než se prst pohnul, ve chvíli předělu mezi nehybností a pohybem, kdy se její záměr stal skutečností. Bylo to, jako když se tříští vlna. Pomyslela si, že kdyby se tak mohla*

---

[44] Poznamenávám, že nerozlišuji pojmy akt, akce, skutek, konání, činnost a jednání. Ve všech případech jde o chování na základě vědomého záměru (což nemusí platit pro libovolný pohyb těla nebo jeho části). V české fenomenologické literatuře je někdy dávána přednost slovu *jednání*.

[45] Úryvkem se obšírněji zabývám v práci [20] (z níž některé části zde přebírám). V úryvku budu o něco níže pokračovat.

ocitnout na hřebenu vlny, mohla by najít tajemství sebe samé, tu část své osoby, která je skutečně za všechno zodpovědná. ([30], s. 36)

Čemu se vlastně Briony podivuje, a to nikoliv poprvé? Ne tomu, že má před očima jakousi *bytost*, ale tomu, jak k oné – navíc na slovo poslušné – bytosti *přišla*! Kdyby si ten masitý pavouk dělal co by chtěl, kdyby se třeba vysmekl a utekl do kouta, Briony by se zřejmě nedivila. Máme však spolu s Briony před očima něco, co není samostatné, co poslouchá – dokonce na „slovo"! *Slovo* zde znamená vědomý záměr, který zpravidla lze verbalizovat, *na slovo* pak navozuje soulad mezi chtěním a provedením, mezi úmyslem a jeho splněním.

Ale přesto ruka má svou vlastní, osobitou roli, například roli uchopovací. Je-li to stroj *na* uchopování, je na mnohé výkony již předem připravena, své dovednosti si nosí s sebou. Ty jsou zčásti nebo úplně za horizontem (chce se říci *pod* horizontem) vědomé kontroly. Chci si třeba zapnout knoflík. Stačí rukám naznačit, *co* mají udělat (třeba: „Zapněte mi prosím knoflík!"), ony pohotově převezmou složitý úkol a provedou jej samy do všech pohybových podrobností (že vlastně ani nevím, jak to dělají, si uvědomím, až když si musím zapnout knoflík jen jednou rukou, levou). Vlastní inteligence ruky se projeví při zapínání knoflíku, zavazování tkaniček, při hře na klavír anebo míchání karet. Ruku lze mnohému naučit, ale něco už musí v sobě mít, co se projeví právě schopností něčemu se takto naučit. Stačí jen pohlédnout na svou vlastní naučenou ruku v akci. (Nesmíme jí do toho příliš mluvit, mohla by nás začít poslouchat a jsouc poslušná – *na slovo* –, zapomněla by na svou šikovnost.)

Briony jeden ze svých prstů ohnula a pak narovnala. *Mystérium tkvělo v okamžiku, než se prst pohnul, ve chvíli předělu mezi nehybností a pohybem, kdy se její záměr stal skutečností.* Zdalipak je ten okamžik předělu postižitelný? Postižitelnost zde ovšem znamená zkušenostně zachytit, zažít, prožít. To však vyžaduje čas: jak víme, bodové okamžiky jsou daleko za (chce se říci hluboko pod) horizontem vědomí. K tomu píše Wittgenstein:

> Konání samo vypadá jako něco, co nemá žádný objem zkušenosti. Jako jakýsi nerozprostraněný bod, špička jehly. Vlastním agens se zdá být právě tato špička. A to, co se děje v jevové skutečnosti, jako by bylo jen následkem tohoto konání. Zdá se, že ‚konám' má zcela určitý smysl, od jakékoli zkušenosti oddělený. ([55], par. 620).

Co je pro Wittgensteina špička jehly, nazývá Briony hřebenem vlny – všimněme si však jistého rozdílu: Wittgensteinova představa je statická (bod, špička jehly), Brionina dynamická – nepostižitelný přechod od hladké vlny k tříšti. Jako by znala teorii nelineárních systémů. Z jiného hlediska to připomíná vnitřní prožitek přítomného času, který, jak víme ze sekce 5, není ostře bodový.

Brionina meditace v dětském pokoji pokračuje:

Příklad 11. (Brionin prst – pokr.)

> *Dala si ukazováček blíž k obličeji a upřeně se na něj zadívala, pobízela ho, aby se pohnul. Zůstal nehybný, protože předstírala, nemyslela to úplně vážně, a protože chtění, aby se pohnul, nebo chystání se s ním pohnout, nebylo totéž, jako s ním skutečně hýbat. A když pak nakonec prst přece pokrčila, jako by ta činnost začala v samotném prstu, ne v některé části její mysli. Kdy poznal, že se má pohnout, kdy*

> *ona poznala, že s ním má hnout? Nepřistihla se při ničem. Bylo to buď anebo.* (tamt.)[46]

Zdálo by se, že Briony si takto empiricky dokázala existenci rozdílu mezi úmyslem (chtěním a chystáním se) a skutkem (pohybem). Zde však musíme být opatrní. Pokud bychom chtěli takovýto rozdíl fenomenálně vykázat i v dokončeném jedinečném aktu (skutečném pohybu prstu), čili rozložit jej explicitně na dvě prožitkově oddělené fáze (úmysl a skutek), nepodařilo by se nám to (na rozdíl od dříve zmíněného rozlišení předchůdné intence Alenky sníst koláček a jeho snědením).

K této otázce stojí za povšimnutí další pasáž z Wittgensteina:

> Ale nezapomínejme na jednu věc: když ‚zvedám ruku', tak se moje ruka zvedá. A vyvstává problém; co je to, co zbude, když od skutečnosti, že zvedám ruku, odečtu to, že se moje ruka zvedá? (tamt., par. 621).

Cožpak lze takto „odečítat" jedno od druhého? Konceptuálně asi ano, šlo by o obecnou otázku, zda úmyslný pohyb je něco víc než událost, než fyzický pohyb, který můžeme pozorovat, ba i zakoušet. Lze však odečítat prožitky jako takové? Pravda, dovedu si představit, že spatřím svou ruku, jak se sama zvedá bez mého vědomého přičinění, podobně jako se zvedá moje noha po úderu kladívka neurologa. To by však byl zcela jiný pocit, než když vidím, jak se moje ruka zvedá právě díky mé snaze ji zvednout. Takže nevím, *jak* a *co* bych měl v druhém případě odečítat.

Z podobného důvodu není zcela jasné, o jakém „zbytku" by se zde vůbec dalo uvažovat: pokud by šlo o pocit z úmyslu jednat bez skutečného jednání, jako když Briony *předstírala, nemyslela to úplně vážně*, pak by šlo o něco jiného než myslet to vážně a odečíst výsledek. A pokud by to měl být vážně míněný pokus zvednout ruku, která by však neuposlechla v důsledku paralýzy či vnější zábrany, případně když by to byl fantom ruky (po amputaci), pak by to opět bylo něco jiného než prožitek úspěšného zvednutí vlastní (zdravé) ruky. Takové případy neuposlechnutí jsou známy v neurologických a psychiatrických kazuistikách a postižený subjekt je zpravidla prožívá s rozpaky, údivem, či frustrací, které lze stěží „odečíst" od celkového prožitku.[47] Naproti tomu „úmyslný pohyb je charakterizován nepřítomností podivování se" (tamt., par. 628), a proto svou vlastní zručnost běžně neobdivujeme.

Ptejme se tedy, zda Brionin experiment s prstem pomáhá objasnit problém formulovaný Wittgensteinem („co zbude, když … odečtu to, že se moje ruka zvedá?") a co by to znamenalo pro pochopení aktuálně hraničního prožitku osobního přispění při uskutečnění úmyslného pohybu. Právě zvedání ruky si volí jako příklad intencionálního aktu J. R. Searle, který svým charakteristicky analytickým stylem rozebírá rozdíl mezi předchůdnou intencí (předsevzetím) něco učinit [*prior intention*] na jedné straně a intencí v aktu [*intention-in-action*], k níž dochází při fyzickém pohybu těla, na straně druhé. Píše:

> Máte-li intenci v aktu [například zvednout ruku], u níž však nedošlo ke splnění podmínek uspokojení [*conditions of satisfaction*; zde: že se vaše ruka zvedla], aspoň jste se pokusili. V typickém případě předem rozmyšleného aktu, při němž jednám na základě předchůdné intence,

[46] Opravený překlad.

[47] Platí to i pro inverzní případ, kdy je uměle vyvolán nechtěný pohyb. Často citován je případ Penfieldova pacienta, jehož motorická oblast mozkové kůry byla elektricky stimulována s následným svalovým pohybem. Pacient reagoval slovy: „Já to neudělal, vy jste to udělal!" (viz [38]).

> jako při zvedání ruky [třeba v úmyslu hlasovat], je struktura celé záležitosti tato: nejdříve jsem vytvořil předchůdnou intenci (jejíž podmínkou uspokojení je, že právě ona má způsobit celý akt) a poté jsem celý akt provedl, kde celý akt má dvě komponenty, intenci v aktu a pohyb těla (přičemž podmínkou uspokojení intence v aktu je, že právě tato intence v aktu má způsobit dotyčný pohyb těla). ([42], s. 45).

Takovýto rozklad umožňuje Searlovi odhalit dvě „mezery" [*gaps*] v kauzálním řetězci, které jsou obě otevřeny – je v nich vůle(!) – pro zásah svobodné vůle. První mezera je mezi rozvažováním (například, zda hlasovat pro nebo proti) a rozhodnutím jakožto předchůdnou intencí (úmyslem hlasovat pro), druhá mezera je mezi předchůdnou intencí (úmyslem hlasovat pro) a intencí v aktu (pokusem zvednout ruku).[48] A Searle poznamenává:

> Mezi intencí v aktu a pohybem těla neexistuje takováto mezera. Pokouším-li se něco opravdu udělat a jsem-li úspěšný, pak můj pokus musí kauzálně stačit k úspěchu. (tamt., s. 50).

Rozumím tomu tak, že pokus (snaha) zvednout ruku již neposkytuje prostor k rozhodnutí ruku nezvednout (což je vlastně analytické tvrzení o pojmu „pokus"). Slova „pokus" a „snaha" jsou zde ovšem zbavena jisté své konotace, podle níž by přitom bylo nutno vynaložit úsilí proti nějakému odporu. Tato konotace vystupuje naopak do popředí v případě paralýzy nebo vnější fyzické zábrany pohybu a vedla též asi Wittgensteina k výroku „Když zvedám ruku, tak se většinou *nepokouším* ji zvednout." (cit.d., par. 622).

Co z toho plyne pro Brionin experiment s prstem? Přijmeme-li Searlovu analýzu, dalo by se říci, že se Briony podařilo fenomenálně (prožitkově) doložit faktický rozdíl mezi úmyslem, jakožto předchůdnou intencí, a pokusem (snahou), jakožto intencí v aktu, nikoliv však rozdíl mezi pokusem (o pohyb) a fyzickým pohybem *per se*. Když Briony prst *pobízela, aby se pohnul*, ale *nemyslela to úplně vážně*, lze to interpretovat tak, že uměle zvětšovala existující mezeru mezi předchůdnou intencí (pohnout prstem) a intencí v aktu (uskutečnit tento pohyb). *A když pak nakonec prst přece pokrčila* – pak už neexistuje žádný prostor k rozhodování, je to již úspěšný akt sám, v němž, jak jsem se pokoušel výše naznačit, není fenomenální rozdíl mezi pokusem učinit pohyb a fyzickým pohybem. Rozdíl je ovšem konceptuální, což vyplývá z existence neúspěšných aktů (patologické případy, vnější zábrany a umělá stimulace – viz výše). Bez konceptuálního rozlišování by ostatně Searle nemohl mnohé ze svých tezí formulovat.

Zdůrazňuji, že řeč byla o *fenomenálním* (1PP) rozlišení – čili jako by šlo o dva různé vědomé prožitky (například oddělené v čase), oba v perspektivě první osoby. Skutečnost, že v 1PP neprožívám odděleně pokus učinit pohyb a pohyb sám, neznamená, že rozdíl se neobjeví v 3PP. Pohyb lze nejen vnímat, ale i registrovat, měřit, vykládat vědecky odkazem na příslušné fyziologické procesy; nic z toho se ovšem nevztahuje k pocitu *jaké to je* pohyb učinit. Takto se na horizontu prožívání (v modalitě pocitu aktéra) dostává perspektiva první osoby (1PP) opět do těsné blízkosti perspektivy třetí osoby (3PP), aniž by však v tomto případě jakkoliv splynuly.

Briony se dále ptá: *Kdy* [prst] *poznal, že se má pohnout…?* Disociace mezi mnou a mým prstem se objeví jakmile bych se snažil – podobně jako Briony – přiznat prstu jeho vlastní vůli. Pak bych totiž nekrčil prst, pouze bych jej o to žádal, a jen tak by mělo smysl jej chválit za to, že „na slovo poslouchá". Nemusel by totiž.

---

[48] Searle zmiňuje ještě třetí mezeru, která se uplatňuje při jednání s časovým rozpětím.

V takovém případě (při vědomém reflexování své akce) by bylo možno odhalit další rozdíl: mezi mým „pokynem“ (intencí v akci), jemuž rozumím, a realizovaným fyzickým pohybem, jemuž nerozumím ve všech detailech provedení, takže nemám jinou možnost, než se spolehnout na prst sám. Jenom on (spolu se svým řídicím centrem kdesi v motorické části mozku), poté, co *poznal, že se má pohnout*, ví, *jak* to zařídit. Umíte stříhat ušima?

11 Pocit vlastní tělesnosti

Pocit, že jsem aktérem v nějakém dění, zejména pokud toto dění spočívá ve fyzickém pohybu nebo je jím zprostředkováno, úzce souvisí s pocitem vlastní tělesnosti [*embodiment*].[49] Jde mimo jiné o tradiční téma filosofů, které se v poslední době dostává do centra pozornosti i v rámci kognitivní vědy (např. [12, 47, 53]). Zde se přidržím již dříve citované nedávné monografie Shauna Gallaghera [12] která je mimo jiné důležitá ve své snaze integrovat fenomenologii (v 1PP) [*first-person phenomenology*] a vědu o vtělené kognici (3PP) [*third-person science of embodied cognition*].[50]

Gallagher vychází z důležitého rozlišení mezi tím, kdy si vědomě všímáme svého těla a kdy je tělo jen na periferii naší pozornosti. Nejnápadnější je toto rozlišení u těch tělesných akcí, které si uvědomujeme ve fázi intencionálního projektu, avšak jejich tělesnou stránku si jako takovou neuvědomujeme (odkazuji na předchozí sekci). V tomto smyslu je tělo (za normálních okolností a jen v určitém stupni) zkušenostně transparentní [*experientally transparent*] (s. 37).

Kde je však přechod mezi zkušenostně transparentním a uvědoměným tělem? Pokud jde o *nehybné* tělo – jeho tvarovou podobu – není to otázka příliš zajímavá. Mé tělo se mi jeví (v 3PP) stejně jako těla druhých lidí; zajímavější jsou ovšem případy tělesných projevů provázejících emoce (nebude to zde mým tématem) a tělesných akcí (úmyslných pohybů), jimiž jsem se zabýval již v předchozí sekci. V příkl. 10 (Brionin prst) jsme si všimli, že pohyb ruky nebo prstu má jednak složku autonomní (*stroj na uchopování*), více méně přístupnou v rámci 3PP, a jednak složku vědomě řízenou subjektem (*záměr se stal skutečností*).

Gallagher věnuje zvláštní pozornost *propriocepci* (smyslu pro pozici různých částí těla na základě údajů ze svalových a kloubových receptorů polohy). Právě u ní lze nalézt přechod od nevědomé, subperzonální *informace* k proprioceptivnímu *uvědomění si svého* těla, které je ovšem v běžném životě nereflektováno, jsouc na periferii či zcela mimo naši pozornost.

Jak už jsem dříve poznamenal (v sekci 4), Gallagher konceptuálně rozlišuje mezi tím, co nazývá *tělesným obrazem* [*body image*] a *tělesným schématem* [*body schema*] (oba termíny užívali již předtím jiní autoři v různých významech, Gallagher se snaží o jejich přesné vymezení). Tělesný obraz je dle něho tvořen „systémem percepcí, postojů a přesvědčení vztahujících se k vlastnímu tělu,“ zatímco tělesné schéma je „systém senzomotorických schopností, které fungují nevědomě či aspoň bez nutnosti uvědoměného monitorování.“ (s. 24). Gallagher varuje před zaměňováním těchto dvou konceptů těla (s. 38). Snadno si tento rozdíl uvědomíme, vzpomeneme-li si na vlastní

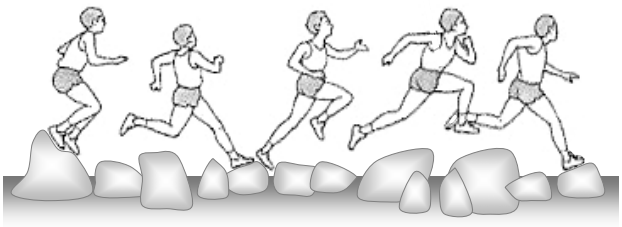
[49] Kdyby existovala telekineze, souvislost by byla asi menší.

[50] Odkazy na stránky v této kapitole se vztahují k této publikaci.

pohyby ve snu: moje snová ruka se zvedá, moje reálná ruka je pod peřinou. Při snění se tělesný obraz účastní snu, zatímco tělesné schéma je vyřazeno.

Nehledě na toto varování soudím, že lze porovnávat, ba i kombinovat zmíněné dva koncepty těla (jeden v 1PP, druhý v 3PP) v konkrétních případech hraniční zkušenosti. Nelze totiž očekávat, že existuje nějaká *ostrá* hranice našeho prožívání těla (v 1PP), za níž by už byl možný jen popis podrobností fyziologických procesů (ve 3PP – dokonce jen ve vědeckém jazyku). Spíše jde o *vlastní horizont tělesného obrazu* (ve smyslu sekce 3), na němž by se 1PP a 3PP přístupy mohly spolu i prolínat.

Jako názorný příklad různých stupňů uvědomění si těla bych rád uvedl jeden typ zkušenosti, kterou většina z nás asi má (nazvu ji *efekt kamzíka*). Představte si, že se snažíte přeběhnout řeku po vyčnívajících kamenech (musíte běžet rychle, abyste udrželi rovnováhu – viz obr. 6). Setkávají se zde tři úrovně pohybu podle stupně vědomé pozornosti: (1) distální intence (chcete se dostat na druhý břeh); (2) bezprostřední (proximální) intence (na který kámen se právě chystáte stoupnout); (3) senzomotorické funkce operující pod hladinou autoreferenční intencionality (není dost času na to, aby se jim vědomí věnovalo, musíte běžet rychle, jako kamzík po skalách).



Obrázek 6. Efekt kamzíka

Jak vidno, je zde horizont tělesného obrazu dynamicky spojen s modalitou vnímání času (sekce 5). Lze to vidět na příkladu (skutečného) kamzíka běžícího po kamenné suti: kameny mizí z jeho zorného pole dřív, než o nich potřebují vědět zadní nohy – jako by se zde uplatňovala retence vjemu. Současně kamzíkův zrak vybírá právě tu a jen tu informaci o tvarech kamenů, která bude až později důležitá pro zadní nohy – protence.

Z tří zmíněných úrovní pohybu je pro nás asi nejzajímavější úroveň (2), jež odpovídá hraničnímu případu v modalitě prožitku vlastní tělesnosti, čili vlastnímu horizontu tohoto prožitku. Úroveň (3) je již za tímto horizontem. Gallagher píše:

> Znalost toho, co dovedou mé nohy, je *v mém těle*, nikoliv v reflektujícím či intelektuálním postoji, který bych ke svým nohám mohl zaujmout. (s. 74).

Smysl pro tělesnost lze dále rozkládat na několik dílčích modalit. Dosud jsem mluvil spíše jen o jedné z nich, *smyslu pro kinestetické řízení pohybu*. Jiná dílčí modalita je *pocit vlastnictví těla* [*sense of body ownership*], čímž se rozumí přirozený předreflexivní pocit, že toto tělo je *moje*, a to včetně – v normálním případě – jeho (tedy *mých*) pohybů. V tom posledním je tento pocit příbuzný, nikoliv však totožný s pocitem *autorství akcí* (pocitem aktéra).

Jestliže Briony medituje nad vlastní rukou (*Jak k této bytosti, tomuto stroji na uchopování, masitému pavoukovi přišla?*), je její ruka sice ve své poslušnosti obdařena

svým svérázným bytím, ale ať už jde o „bytí“ strojní (algoritmické) nebo pavoučí (zčásti samostatné), vždy je něčím, co je *její,* Brionino. K čemu *přišla*, to je její.

Pocit vlastnictví předchází každý akt upření pozornosti k vlastnímu tělu, a proto v normálních případech nemá smysl říkat, že si tělo přivlastňuji – musel bych si ho totiž nejdříve “odvlastnit”, což je postoj, který najdeme snad jen u určitých patologických případů[51]. Ostatně jsou to právě takové případy (viz zmínka v předchozí sekci), které by mohly odlišit pocit vlastnictví od pocitu autorství. Pacient by mohl tvrdit: „Je to *váš* pohyb, pane doktore, i když se pohybovala *moje* ruka. Nebyl to pohyb *můj* – mnou chtěný a iniciovaný.“ (viz pozn. 46). V normálních případech se nicméně cítíme vlastníky i nechtěných, bezděčných a samovolných pohybů, aspoň se za ně často cítíme odpovědni: „Omlouvám se, že jsem vám šlápnul na nohu.“

Intuitivním horizontem tělesnosti je ovšem i prostorová hranice fyzického těla. Je to typ horizontu vlastního, který je v každé přirozené situaci, k níž náleží vtělený subjekt. I u virtuálního subjektu si (již ne tak samozřejmě) představujeme jeho virtuální *tělo*. Na jedné straně, v diskursu 3PP, je hranice těla více méně záležitostí té či oné vědní disciplíny (možná v různé podobě – např. imunologie vs. etologie) a může mít smysl i pro jiné organismy, než je člověk. A. Damasio poznamenává:

> [Hranice těla] u složitých organismů, jako jsme my, může mít různé podoby – například pokožka, pokrývající většinu našeho těla, rohovka, pokrývající část oka, sliznice v ústech. Bez hranice není těla, a bez těla není organismu. Život potřebuje hranici. ([7], s. 137)

Na druhé straně je uvědomování si hranic vlastního těla (v 1PP) závislé na (nejen emočním) postoji k němu a je součástí tělesného obrazu. V mnoha ohledech nemůže toto uvědomování soutěžit s 3PP popisem: nevidím rohovku svého oka ani sliznici v mých ústech, nemluvě o vnitřních funkcích, orgánech a buněčné tkáni.

V jiných ohledech si lze myslet hranici „těla“ rozšířenu daleko za hranice fyzického těla. Zdalipak by se dříve citované Merleau-Pontyho diktum „moje tělo je kdekoliv tam, kde je třeba něco vykonat“ dalo interpretovat tak, že horizont mého těla objímá celek pole mé aktivity? S dávkou libovůle bychom v takovém případě mohli mluvit o distálním horizontu modality tělesnosti.

12 Vědomí sebe

Úvahy o pocitu aktéra (na rozdíl od smyslu pro působení obecně) a vlastní tělesnosti (na rozdíl od jiných těles) nás natolik přiblížily k vědomí *sebe*, že se nelze vyhnout tematizaci tohoto typu vědomí jako specifické modality prožívání situací. Čím závažnější je to téma, tím obtížnější je o něm – jako o tématu – mluvit. Omezím se zde jen na některé jeho aspekty, zejména pokud souvisejí s neodmyslitelnou rolí subjektu v epizodické situaci (či přímo subjektu situace).[52]

Připomínám, že když hovořím o epizodické situaci, činím tak vždy již s implicitní nebo explicitní představou, že existuje *subjekt situace* – osoba, která situaci vnímá, prožívá, ovlivňuje (a případně i – jako v našich příkladech – vypráví). Při zdůrazňování

[51] Existuje o nich bohatá odborná literatura.

[52] Problémem Já a „jáskosti“ se z fenomenologického hlediska a v podobě přístupné i pro kognitivní vědu zabývá Zahavi [56, 57].

pohledu 3PP též lze hovořit o *protagonistovi* (hlavní osobě) situace, který tudíž k situaci *patří*. Ale i naopak, situace *patří* subjektu, je *jeho*. Při reflexi vlastních situací tak či onak oscilujeme mezi pohledy 3PP a 1PP a proto i v textu této studie často přecházím ke (gramatické) první osobě, jako bych to byl *já*,[53] jako by to byla *moje* situace. Kdykoliv situaci sám prožívám, ať již aktuálně, ve vzpomínce, anebo empaticky se vžívaje do situace druhého, vždy jsem v ní jaksi přítomen. Jinak si lze stěží představit reflexi s tendencí k 1PP.

Vědomí sebe, *jáskost* [*selfhood*], je předreflexivně, inherentně a periferálně přítomno v každém kognitivním aktu, pozorování, zájmu, myšlence, tělesném pohybu, chtění. Tím spíše i v reflexi těchto aktů: Briony (z příkl. 10 a 11) si nakonec – až poté, co pohnula prstem – uvědomuje, že za touto akcí bylo její vlastní „já“:

Příklad 12. (Brionin prst – další pokr.)

> *Nebylo zde žádné viditelné propojení, a přesto je to skutečně její já, to ví, pod tou hladkou, ničím nenarušenou tkaninou – je to její duše? –, co rozhodlo, že zanechá předstírání, a dalo konečný příkaz.* ([30], s. 36)[54]

Podobně i John Searle, když zkoumá strukturu intencionálního jednání, dospívá k představě jáskosti, která „přichází sama od sebe“ u jakékoliv „vědomé intencionální bytosti schopné se angažovat ve svobodném jednání, podloženém důvody.“ ([42], s. 145).

Přejít od předreflexivní „přítomnosti“ vlastního Já[55] k jeho reflexi v sobě skrývá závažný problém. Obrátím-li pozornost k sobě samému, znamená to, že ze svého Já činím *objekt* reflexe? Ale tím by se mi vymklo právě to pravé Já, k němuž jsem se chtěl dostat. Místo něho přede mnou vyvstane nový horizont, který nelze žádnou reflexí překročit a za nímž se skrývá jakési unikavé, transcendentní Já. Cituji Jana Patočku:

> Prožívání je jakoby tkáň napjatá mezi dvěma horizonty: jedním je mé já, druhým je svět. Prožívání je jakási explicitace těchto horizontů, s tou zvláštností, že abych mohl explikovat sebe, musím nejprve stanout na půdě světa a pak teprve se vrátit k sobě. [36]

Schematické znázornění horizontu z obr. 1 si můžeme překreslit podle obr. 7, kde **H'** znázorňuje horizont reflektovaného „já“ a **S'** je mé druhé, reflektující Já.[56] Hledím totiž odněkud „vně“ situace, o níž mi v reflexi jde (být „na půdě světa“ zde neznamená být uvnitř situace).

Damasio [7] rozlišuje v diskursu 3PP dva druhy Já, a to odvozeně od dvou druhů vědomí, *jaderného* [*core consciousness*] a *rozšířeného* [*extended consciousness*] ([7], s. 17). *Jadernému Já,* které je krátkodobé a pomíjivé, by v našem pojetí odpovídala modalita jáskosti pro atomické či aspoň nepříliš rozsáhlé epizodické situace, kdežto dlouhodobé, *autobiografické Já* by bylo odvozeným pojmem, vztaženým k celoživotnímu panoramatu (viz sekce 2). Pro naše účely by pak bylo vhodné mluvit i o mezilehlém, *epizodickém Já*, které by odpovídalo nejen krátkým, ale i delším epizodám (s jakými jsme se setkali

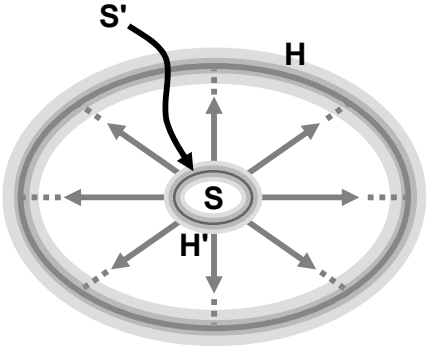
[53] Někdy i „ty“ – viz konec pozn. 10.

[54] Opravil jsem překlad.

[55] Iniciála dává slovu „Já“ význam abstraktního singuláru (*to* já).

[56] Husserl v tomto smyslu odlišoval objektivizované, konstituované, *transcendentní* Já od konstituujícího, *transcendentálního* Já ([56], s. 183).

v našich příkladech). Epizodické Já by si „samo sebe“ uvědomovalo s různou intenzitou, proměnnou i v průběhu prožívané epizody (tak když v příkl. 8 Alenka *dostala šťastný nápad,* její pozornost se tím přesunula z koláčků na *její vlastní velikost* a tím implicitně i na její epizodické Já).



Obrázek 7. Problém sebereflexe (**S** – reflektovaný subjekt, **S'** – reflektující subjekt, **H'** – horizont sebereflexe).

Možnost různé intenzity prožitku sebe sama nastoluje otázku prahového vědomí sebe. Jeden směr hledání by mohl být prožitek při probouzení. Často je citována pasáž ze začátku „Hledání ztraceného času“, kde Proust líčí průběh probouzení, jako by hybridní paměť (abych použil terminologii ze sekce 2) „jako lano visící z nebe“ vytahovala probouzející se Já z propasti nebytí. Damasio popisuje probouzení takto:

> Objevují se obrazy lidí, věcí a okolní scény a přesto, po jistou krátkou dobu, která se nám však může zdát dost dlouhá, chybí vědomí sebe, stejně jako pocit vlastního myšlení. O zlomek sekundy později je smysl pro Já „zapnut“ – a opravdu, vágně se nám zdá, že obrazy jsou naše, i když všechny podrobnosti ještě do sebe správně nezapadají. Poněkud déle trvá, než se autobiografické Já obnoví jako průběžné dění a než se situace stane zcela srozumitelnou. ([7], s. 94)

Souhlasím s představou zpožděného obnovování autobiografického Já, nejsem si však jist, zda se dá o jeho chybění takto mluvit – byla by to totiž náhlá změna diskursu z 1PP do 3PP. Podle předchozích úvah lze stěží prožívat sebe (v 1PP) jako něco, co je buď „zapnuto“ nebo „vypnuto“; spíš snad jen jako něco, co je více nebo méně prožíváno – čili více nebo méně v centru našeho uvědomění (a reflexe).

V reflexi si snad mohu představit absenci vlastního těla, nikoliv však absenci svého Já *per se*. Já je vždy přítomno, i když (dokonce většinu času) jen okrajově. „Přítomnost“, stejně jako „absenci“ ovšem chápu jen ve vztahu k *vědomé* zkušenosti; o komatu (ani o bezesném spánku) neexistuje aktuální zkušenost v 1PP – o takových stavech mluvíme jen z odstupu, tedy v jazyku 3PP.

Pokud se nemýlím v tom, že smysl pro vlastní Já je vždy nějak „přítomen“ a současně že toto Já nemá povahu objektu, protože je vždy skryto za horizontem sebereflexe, pak bychom se vlastně neměli divit, že je prožíváme jako identické samo se sebou. Mluvím o normálním prožívání, nikoliv patologickém – s takovým (naštěstí) nemám vlastní zkušenost.

S tím souvisí i to, že i ve snech mám pocit, že jsem to stále já, nezávisle na tom, že mohu mít jiné vlastnosti, jinou (snovou) minulost, jiný vzhled:

Příklad 13 (Můj sen o černochovi[57])

> *Chystám se na raut nějaké vědecké konference, ale mám zpoždění. Mám se převléci? Dívám se do zrcadla: jsem statný zavalitý černoch ve žlutém tričku a jeansech. Uvažuji: mohl bych jít, tak jak jsem, ale někdo by mohl být na nevhodně oblečeného černocha hrubý. Snad by pomohlo, kdyby moje tvář měla podobu předsedy konference. A hle, už tuto podobu má. Epizoda končí.*

Ve snu mohu měnit podobu, mohu být i černochem, ani okolní situace, ani moje minulost nemusí nijak souviset s mým bdělým životem – avšak jedno vím zcela neproblematicky: jsem to stále *já*, který svou snovou situaci prožívám, ba i který se shlížím ve sněném zrcadle. Ani se nemusím spoléhat na svou autobiografickou paměť, aby mi zaručila identitu.

Totéž konstatuje známý švýcarský psychiatr a psychoterapeut, zakladatel daseinsanalýzy Medard Boss:

> Ve snu se ovšem někdy mohu stát jiným člověkem nebo zvířetem, které právě šlo vedle mě. Ale i v každé snové proměně to zůstávám „já“, kdo se proměňuje, a zůstávám stále týž. ([3], s. 145)

13 Smysl pro druhého

V epizodické situaci se mohou samozřejmě vyskytovat i jiné osoby, než je subjekt situace. Bylo tomu tak v příkl. 7, kde se kromě Briony nacházela i Cecilia, Robbie a stařena na chodníku. Tyto další osoby mohou být pro subjekt ve větší nebo menší míře důležité a také mohou být více či méně do situace zataženy jako postavy děje. Vztah k druhým osobám je pro člověka natolik důležitý, že stojí za to o něm pojednat jako o samostatné modalitě prožívání – *smyslu pro druhého* [*otherness*].

Přečtěte si následující úryvek z románu M. Kundery; subjektem situace (a hlavní osobou celého románu) je žena jménem Irena:

Příklad 14 (Irena)

> *Na protější lavici uviděla jakéhosi muže a po několika okamžicích nejistoty a údivu ho poznala. Vzrušeně čekala, až se setkají očima, a pak se usmála. Odpověděl také úsměvem a lehce pokývnul hlavou. Vstala a vykročila k němu, on zase vykročil k ní. “Neznali jsme se v Praze?” řekla česky, “pamatujete si mě?”* ([29], s. 46)[58]

Popsána je epizodická situace, během níž se Irena setkává s druhou osobou, která se během krátké doby postupně promění z neznámého cizince v dávného přítele. Děj (zápletka) sestává z několika fází (podepizod), které po sobě následují v objektivním (ovšem fikčním) čase: *Irena vidí muže* → *chvíle nejistoty* → *poznává ho* → *čeká na pohled* → *úsměv* → *jeho úsměv* → *Irena vstává* → *kráčí k němu* → *on vstává* → *Irena promluví.* Prosté zahlédnutí (*jakéhosi muže*) by pro Irenu nic nemuselo znamenat, pokud by u ní nevznikla nejistota (*že bych ho znala?*), pak údiv (*musím ho znát!*) a rozpoznání (*ano, znám ho!*). Následuje důležitý okamžik, kdy Irena vstoupí v předverbální komu-

[57] V noci na 9.8.2006.

[58] Můj překlad.

nikaci s oním mužem (*výměna pohledů, pak úsměvů*); epizoda (pro nás) končí zahájením verbální komunikace.

V rámci fenomenologicky orientované kognitivní vědy se vztahu k Druhému[59] věnuje Dan Zahavi, který rozlišuje čtyři různé teze ([57], kap. 6):

(1) Přímé setkání s druhým je provázeno specifickým naladěním vědomí, zvaným *empatie*;

(2) Setkání s druhým (včetně empatie) je podmíněno jistým typem *jinakosti vtěleného Já* [*alterity of the embodied Self*];

(3) Intersubjektivita není redukovatelná na přímé setkání dvou jednotlivců, nýbrž je založena v apriorním vztahu mezi subjektivitou a světem;

(4) Rozhodujícím a nezanedbatelným aspektem intersubjektivity je konfrontace s radikální jinakostí Druhého.

Zahavi navrhuje, aby čtyři možné přístupy podle uvedených tezí nebyly brány izolovaně, nýbrž ve vzájemném propojení (a překonání jejich odlišnosti) v rámci „multidimenzionální teorie intersubjektivity“ ([57], s. 174–177). Ježto nám jde o „konkrétní fenomenologii“ singulárních epizodických situací, můžeme se omezit na první dva body.

U předverbální komunikace Ireny s mužem na protější lavici v uvedené epizodě můžeme – v souhlasu s tezí (1) – předpokládat velkou míru empatie na straně Ireny. Díky ní mohla očekávat reakci partnera a spoléhat na ni. Lze se navíc pokusit tuto empatii přeložit do jazyka 3PP a interpretovat ji jako pohotové užití folkové (či naivní) psychologie.[60] Hypoteticky mohla Irena třeba uvažovat takto:

Příklad 15 (Irena – jejíma očima**)**

> *Musel mě poznat, jinak by neprovázel svůj úsměv pokývnutím hlavy. Bylo to jen lehké pokývnutí, není si zcela jist, třeba si (tak jako já) přesně nevzpomíná na okolnosti našeho dřívějšího setkání. Každopádně ho nepřivedu do rozpaků, když k němu půjdu a zeptám se.*

Je to vlastně *moje* folková psychologie o Irenině folkové psychologii. Kdoví, co jí *opravdu* proběhlo hlavou, dokonce pochybuji, že by situaci takto racionalizovala. Prostě ji tak mohla takto nějak intuitivně „vidět“.

Pro předverbální (občas i pro verbální) komunikaci je setkání pohledů důležitým výrazovým prostředkem. Obousměrnost přitom nelze vyjadřovat jako nějakou výměnu informací. John Ziman píše:

> V běžném použití slovo „komunikovat“ znamená víc, než „informovat“. Není to jen záležitost vyslání zprávy od jedné osoby k druhé nějakým způsobem a za nějakým účelem. Pro Ego komunikovat s Druhým vyžaduje obousměrné přenosy a transakce, při nichž každá strana zpětně reaguje, jakmile pochopí smysl toho, co přijala. ([58], s. 20)

---

[59] V místech, kde je to důležité, užívám iniciálu ve slovu „Druhý“ (podobně jako u „Já“) ke zdůraznění jeho významu jako abstraktního singuláru (*ten* druhý).

[60] Srov. např. [43], kap. 7.

Neverbální komunikace předpokládá určité bezprostřední porozumění jazyku těla, přičemž moje vzorce chování jsou srovnatelné, ba dokonce identické se vzorci chování Druhého. Tím se vlastně dostáváme k druhé uvedené tezi (jinakost vtěleného Já). Zcela podle Merleau-Pontyho výroku, že „bez mé exteriority by druzí neměli interioritu," Irena zná stavy mysli druhých, protože zná sebe – své tělo – jakoby zvenku, v 3PP. Musí rozumět *svému* úsměvu jako stejnému typu stavu jako je *jeho* úsměv (totéž pro její a jeho pohled a další výrazy tváře nebo těla). Ilustruje to Merleau-Pontyho důraz na *tělesnost* [*embodiment*] jedince (spolu s jeho *zasazením* [*embedding*] do prostředí) jako něco podstatného pro existenci mysli.

Takto se u modality smyslu pro druhého smazává rozdíl mezi 1PP a 3PP. Moje prožívané tělo nelze separovat od mého vnímajícího těla a stejně já vnímám Tvoje tělo jako neseparovatelné od Tebou prožívaného těla. Můj postoj k Tobě není ani pravá 3PP (jako k jiným objektům), ani pravá 1PP (jsi pro mne přece jen ten druhý). Tomuto dilematu se vyhneme postulováním *perspektivy druhé osoby* (2PP). Důležitou komponentou 2PP je rozumění expresivním výrazům tváře a těla Druhého – jeho úsměvům, pohledům, pokývnutím, zachmuřením – a ovšem i rozumění řeči.

Je třeba poznamenat, že smysl pro Druhého se vztahuje i k osobám, které nejsou v dané epizodické situaci fyzicky přítomny. Konec konců, jak Merleau-Ponty [32] zdůrazňuje, i prožívání vlastní subjektivity může (či musí) obsahovat anticipaci existence Druhého. Vzpomeňme, že v příkl. 13 (Můj sen o černochovi) se vidím očima nepřítomného diváka (který *by mohl být na nevhodně oblečeného černocha hrubý*). Proč bych se jinak na sebe díval do zrcadla?

Nedotkl jsem se dosud otázky hraničního prožívání druhých. Mohlo by se to vztahovat k případu, kdy přítomná osoba je právě na horizontu zájmu ze strany subjektu situace. S takovým rozlišením jsme se setkali v příkl. 14 (Irena): muž na protější lavici nejprve náležel do podprahové sféry Irenina zájmu o jiné přítomné bytosti (na protější lavici mohlo být více osob, které ani nejsou zmíněny), teprve v průběhu epizody onen muž vystoupil před horizont jejího zájmu, a to výrazně.

V této souvislosti stojí za zmínku prahový či spíše podprahový smysl pro druhé, dokumentovaný Kannerem [26] u šestiletého autistického chlapce (výjimečně uvádím patologický případ, zdá se mi ilustrativní, i když jej nelze popsat v 1PP):

> Nejnápadnější rys jeho chování byl rozdíl v jeho reakcích na předměty a na lidi. Předměty jej snadno pohlcovaly, věnoval jim pozornost a soustavně si s nimi hrál. Zato lidi zjevně považoval za nežádoucí a věnoval jim tak malou pozornost, jakou jen byli ochotni připustit.[61]

14 Socialita

Kolektivita či socialita je zpravidla tematizována jako varianta, či zobecnění smyslu pro druhé. V poněkud jiném směru uvažování však můžeme považovat socialitu za specifickou modalitu prožívání situací. Stejně jako doposud budu ovšem potlačovat mimosituační kontext.

Představte si epizodickou situaci, v níž – jako její subjekt – náležíte k určité skupině či komunitě více osob, které s vámi typicky něco sdílejí (intence, zájmy, názory, rituály apod.) nebo na nichž vám nějak záleží. Můžete zaujímat perspektivu druhé osoby

[61] Citováno Zahavim [57], s. 189–190.

(2PP) odděleně vůči jednotlivým příslušníkům skupiny, to však nestačí k vašemu prožívání své účasti v ní. Je tu něco víc – především máte vztah ke skupině *jako celku*, jako k entitě vyššího řádu, a navíc má tento vztah specifickou podobu: jste jejím *členem*. Na takovém vztahu můžeme založit určité pojetí *sociality* jako další modality prožívání situací. (Pro jednoduchost předpokládejme, že dotyčná skupina je v dané situaci přítomna celá, i když může v jejím průběhu vznikat nebo zanikat – pak by bylo možno mluvit o *epizodické skupině*).

Je myslitelné spojit skupinu se speciálním typem perspektivy – s kolektivním „my“, čili s *perspektivou první osoby plurálu* (1PPP). Smysl pro socialitu tedy spočívá ve dvou na sebe navazujících perspektivách: 1PP člena skupiny ke skupině a 1PPP skupiny jako celku k situaci (ke světu). Pak tím, co pro smysl pro Druhého bylo setkání tváří v tvář, je pro socialitu aktuální náležení do skupiny. Existuje tu také nějaká obdoba empatie? Tuto roli by nejspíše mohl hrát specifický modus vědomí, totiž *identifikace* se skupinou (jako entitou vyššího řádu).

Vhodná chvíle pro příklad; bude opět z McEwanova románu „Pokání“. Skupina infanteristů se před očima Turnera (hlavní postavy úryvku) právě chystá lynčovat letce RAF:

Příklad 16 (Turner)

> *Muži v kruhu se začali pochechtávat. U barového pultu vytušili, že se chystá něco šťavnatého, takže se začali trousit další vojáci. Jak se dav vzdouval kolem kruhu, jakýkoli zbývající smysl pro osobní odpovědnost vymizel. Všech se začínala zmocňovat chvástavá bezohlednost. Někdo tipnul nešťastníkovi cigaretu o hlavu a zvedla se vlna souhlasného hulákání. Směšně vyjekl a to je rozchechtalo. Nenáviděli ho a zaslouží si všechno, co ho nemine.*
>
> […] *Bylo šílenství jít tomu člověku na pomoc, bylo ohavné tak neučinit. Turner současně rozuměl jak rozjařenosti vojáků, tak zákeřnosti, s jakou by se ho mohla rovněž zmocnit.* ([30], s. 219)[62]

V takovéto napjaté situaci je socialita zvlášť dominantní modalitou (pro členy skupiny) ve srovnání s ostatními modalitami prožívání. Skupina se v průběhu epizody aktuálně tvoří (*začali se trousit další vojáci*) a stává se prostředím pro samovolně se amplifikující jevy, typické pro kolektivní systémy (*všech se začínala zmocňovat chvástavá bezohlednost; zvedla se vlna souhlasného hulákání; chechtot; nenávist*). Sám Turner není členem skupiny, nicméně dobře rozumí euforii davu a současně se brání tomu, aby se přidal.

Všimněme si, že uvnitř skupiny se ztrácí pocit individuální odpovědnosti (*smysl pro osobní odpovědnost vymizel*). Zdá se to být v rozporu s tím, co jsem psal v sekci 7 – že pocit odpovědnosti klesá se vzdáleností. To však jen zdůrazňuje specifiku sociality: je-li ona dominantní, může se význam vzdálenosti obrátit. Čím větší a sevřenější je skupina, tím se její členové mohou cítit méně zodpovědni za kolektivní chování.

Je mnoho případů, kdy lze sledovat stupeň identifikace se skupinou (bitvy, hry, vzpoury, rituální praktiky, skandování, sborový zpěv atd.). Searle se v práci [41] zabývá kolektivní intencionalitou [*We-intentionality*] a argumentuje proti její redukovatelnosti na

[62] Opravený překlad.

individuální intencionalitu [*I-intentionality*]. Zde se omezím na zmínku o několika typech hraničního prožívání sociality.

Je-li subjekt členem epizodické skupiny, nemusí mu být vždy jasné, kdo další je a kdo není jejím členem, jinými slovy, jaký je rozsah příslušného „My“ oproti „Oni“. Může existovat odstupňované členství, a to jinak při pohledu v 1PP a 3PP. Druhá možnost: samotnému subjektu nemusí být jasné, zda se sám má či nemá považovat za člena. (Obě možnosti lze charakterizovat určitým typem horizontu.)

Zajímavější hraniční případ odpovídá stupni identifikace subjektu se skupinou, který se může pohybovat mezi dvěma imaginárními extrémy, absolutní identifikací a subverzivním postojem. Prvému extrému by musela odpovídat úplná absence smyslu pro Já (jen zdánlivě je to v rozporu s tvrzeními v předchozí sekci – zde jde totiž o abstraktní, nedosažitelný extrém), druhému by naopak odpovídal velmi výrazný smysl pro Já.

15 Shrnutí a výhledy

V této studii jsem se pokusil otevřít jistý specifický směr uvažování o lidském vědomí. Jako východisko jsem zvolil skutečnost, že v běžném životě se člověk nesetkává s jednotlivými fenomény vytrženými z prostředí, nýbrž že prožívá komplexní situace či životní epizody v jejich celistvosti, omezené jen jakýmsi (pomyslným) horizontem, nepřekročitelným „zevnitř“ situace (v 1PP), avšak v jisté míře přístupným objektivnímu (případně i vědeckému) pohledu (v 3PP). Z toho jsou odvozeny i dvě tendence v reflexivní analýze epizodických situací, u nichž se zpravidla zdůrazňuje jejich vzájemná neslučitelnost (existence tzv. explanační mezery).

Ukazuje se, že v blízkosti horizontu prožívání nemusí jít o něco tak radikálního. Jde tu ovšem o horizont v mnoha odlišných směrech, a proto (a i z jiných důvodů) jsem zvolil metodu analýzy „per partes“ – pojmově rozlišit několik různých modalit prožívání a v první fázi studia se zaměřit na každou z nich odděleně. Jde o otevřený projekt a před námi se ihned rýsuje množství témat, která by si zasloužila další zkoumání, některé přímo vyplynuly z našich úvah. Uzavírám proto několika náměty pro další studium:

(1) Zabývat se druhou fází metody „per partes“, čili ukázat, jak některé (nebo všechny) modality na sobě vzájemně závisí, jak jsou propojeny (některé souvislosti jsem již uvedl na různých místech výkladu).

(2) Zkoumat vztah různých perspektiv: 1PP ↔ 2PP ↔ 3PP, a to v singuláru i plurálu. Jsou všechny stejně původní a stejně přirozené?

(3) Podrobněji než doposud se zaměřit na hraniční (prahové) prožitky, klasifikovat je podle typů a podle případných důsledků pro vztah mezi 1PP a 3PP.

(4) Studovat různé patologické kazuistiky, které by mohly podpořit i některá teoretická rozlišení (v tomto směru lze čerpat z bohaté literatury v oblasti klinické neuropsychologie a psychiatrie).

(5) Rozšířit studium na vztahy mezi různými epizodickými situacemi v rámci životního panoramatu téže osoby, jak se vzájemně ovlivňují, jakou roli hraje identita prvků a identita subjektu.

(6) Zkoumat další příklady epizodických situací ve fikčních světech, v imaginaci, ve snech a ve změněných stavech vědomí. V čem spočívá rozdíl od přirozených epizodických situací.

Poděkování: Tento příspěvek vznikl v rámci řešení výzkumného záměru CTS MSM 0021620845

Literatura

[1] Adler, H.G.: *Panoráma.* Brno: Barrister & Principal 2003.
[1a] Ajvaz, M.: *Prázdné ulice*. Brno: Petrov 2004.
[2] Bergson, H.: *Hmota a paměť*. Praha: OIKOYMENH 2003.
[3] Boss, M.: *Včera v noci se mi zdálo*. Praha: Grada 1994.
[4] Carroll, L: *Alenka v kraji divů a za zrcadlem.* Praha: Albatros 1865/1985 (překl. Aloys a Hana Skoumalovi, 1961).
[5] Cutting, J. E., Vishton, P. M.: Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth. In: *Handbook of perception and cognition,* Vol 5: *Perception of space and motion*. San Diego, CA: Academic Press 1995, pp. 69–117. [http://people.psych.cornell.edu/~jec7/pubs/78.pdf]
[6] Dainton, B.: *Stream of Consciousness: Unity and Continuity in Conscious Experience*. London: Routledge 2000.
[7] Damasio, A.: *The Feeling of What Happens*. San Diego: Harcourt 1999.
[8] Doležel, L.: *Heterocosmica – fikce a možné světy*. Praha: Karolinum 2003.
[9] Edwards, A. B.: The Phantom Coach. In: *The Supernatural Omnibus*, Twickenham, UK: Senate 1864/1998.
[10] Gallagher, S.: *The Inordinance of Time*. Evanston, Ill.: Northwestern University Press 1998.
[11] Gallagher, S.: Od fenomenologie ke kognitivní vědě. *Aluze* **03** (2003) 92–102.
[12] Gallagher, S.: *How the Body Shapes the Mind.* Oxford: Oxford University Press 2005.
[13] Gibson, J. J.: *The Ecological Approach to Visual Perception*. Boston: HoughtonMiffin 1979.
[14] Gibson, J. J.: A Theory of Direct Visual Perception. In: *Vision and Mind: Selected Readings in the Philosophy of Perception*, (A. Noë, E. Thompson, eds.) Cambridge, MA: MIT Press 2002, pp. 77–90.
[15] Havel, I. M.: Prostor k pohybu. *Vesmír* **82** (2003) č. 5, 243.
[16] Havel, I. M.: Prostory nejen k pohybu. *Vesmír* **82** (2003) č. 11, 603.
[17] Havel, I. M.: Modality subjektivně prožívaných situací. In: *Kognice a umělý život IV*. (J. Kelemen a V. Kvasnička, eds.), Opava: Slezská universita v Opavě 2004, 159–178.
[18] Havel, I. M: Čas ze vzpomínek. *Přítomnost* (léto, 2004), 16–18.
[19] Havel, I. M.: A Structure of Experienced Time. In: *Endophysics, Time, Quantum and the Subjective* (R. Buccheri et al.. eds.) World Scientific Publishing Co. 2005, 163–188.
[20] Havel, I. M.: Brionin prst. In: *Viva Activa, Viva Contemplativa – Janu Sokolovi k sedmdesátým narozeninám* (J. Kružík, ed.), Praha: UK FHS 2006, 334–346.

[21] Havel, I. M.: Prožívání epizodických situací – modality a hraniční případy, In: *Kognice a umělý život VI.* (J. Kelemen, V. Kvasnička, eds.), Opava: Slezská universita v Opavě 2006, 167–180.
[22] Havel, I. M.: Experience of Episodic Situations (First draft), Res. Report CTS-06-17 (October 2006).
[23] Havel, I. M., Mitášová, M.: Prostor prožívaný jako prostor k jednání. In: *Prostor a jeho člověk* (M. Ajvaz, I. M. Havel, M. Mitášová, eds.), Praha: Nakl.Vesmír 2004.
[24] Husserl, E.: *Přednášky k fenomenologii vnitřního časového vědomí.* Praha: Ježek 1917/1996.
[25] James, W.: *The Principles of Psychology.* New York: Dover 1950/1890.
[26] Kanner, L.: Autistic disturbances of affective contact. *Nervous Child* **2** (1943) 217-250.
[27] Koch, C.: *The Quest for Consciousness: A Neurobiological Approach*. Englewood, CO: Roberts and Co. 2004.
[28] Koukolík, F.: *Lidský mozek*. Praha: Portál 2000.
[29] Kundera, M.: *Ignorance* (překl. do angl. L. Asher). London: Faber and Faber 2002.
[30] McEwan, I.: *Pokání* (překl. M. Válková). Praha: Odeon, 2003.
[31] McTaggart, J. M. E.: The unreality of time. *Mind* **17** (1908) 457–474.
[32] Merleau-Ponty, M.: *Phenomenology of Perception*, (překl. C. Smith) London: Routledge 1945/1962.
[33] Michotte, A. (1946): *The perception of causality* (překl. T. R. Miles & E. Miles). New York: Basic Books 1963.
[34] Němec, J.: Fenomenologická metoda. *Souvislosti* (2001/3–4), 120–121.
[35] Noë, A.: *Action in Perception*, Cambridge, MA: The MIT Press 2005.
[36] Patočka, J.: Koncept přednášky o tělesnosti. (rukopis 1980/8), In: *Přirozený svět a pohyb lidské existence* (I. Chvatík, ed.) Archivní soubor, Praha 1980.
[37] Patočka, J.: *Přirozený svět jako filosofický problém*. Praha: Československý spisovatel 1936/1992.
[38] Penfield, W., Rassmusen, T. B.: *The Cerebral Cortex of Man*. New York: Maxmillan Books 1950.
[39] Poe, E. A.: *Zánik domu Usherova.* Praha: Krásná kniha 1927.
[40] Popper, K. R.: *The Open Universe: An Argument for Indeterminism*. London: Routledge 1988.
[41] Searle, J. R., Collective intentions and actions. In: *Intentions in Communication* (P. Cohen et al., eds.) Cambridge, Mass.: The MIT Press 1990.
[42] Searle, J. R.: *Rationality in Action*, Cambridge, MA: The MIT Press 2001.
[43] Sedláková, M.: *Vybrané kapitoly z kognitivní psychologie*. Praha: Grada 2004.
[44] Shallice, T.: The detection of change and the perceptual moment hypothesis. *Br. J. Stat. Psychol*. **17** (1964) 113–135.
[45] Sokol, J.: *Rytmus a čas*. Praha: OIKOYMENH 1996.
[46] Stern, D. N.: *The Present Moment in Psychotherapy and Everyday Life*. New York: Norton 2004.
[47] Bermúdez J. L., Marcel, AS., Eilan, N. (eds.): *The Body and the Self.* Cambridge, MA: The MIT Press 1995.

[48] Tulving, E.: Episodic and semantic memory. In: *Organization of Memory* (E. Tulving & W. Donaldson, Eds.). New York, NY: Academic Press 1972, 382–402. [http://individual.utoronto.ca/akim/12.EpSem72.pdf]
[49] Tulving, E.: Episodic Memory: From Mind to Brain. *Annual Review of Psychology* **53** (2002) 1-25.
[50] van Gelder: Wooden iron? Husserlian phenomenology meets cognitive science. In J. Petitot et al. (eds.) *Naturalizing Phenomenology*. Stanford, CA: Stanford University Press 1999, pp. 245-265.
[51] van Peursen, C. A.: The Horizon. In: *Husserl: Expositions and Appraisals* (A. Elliston and P. McCormick, Eds.) Notre Dame, Indiana: Univ. Notre Dame Press 1977, pp. 182–200.
[52] Varela, F.: The specious present: A neurophenomenology of time consciousness. In J. Petitot et al. (eds.) *Naturalizing Phenomenology*. Stanford, CA: Stanford University Press 1999, pp. 266–314.
[53] Varela, F. J., Thompson, E., Rosch, E.: *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge, MA: The MIT Press 1991.
[54] Vopěnka, P.: *Úvod do matematiky v alternatívnej teórii množin.* Bratislava: Alfa 1989.
[54a] Vopěnka, P.: *Meditace o základech vědy*. Praha: Práh 2001.
[55] Wittgenstein, L.: *Filosofická zkoumání*. Praha: Filos. ústav AV 1993 (překl. J. Pechar).
[56] Zahavi, D.: *Self-Awareness and Alterity: A Phenomenological Investigation.* Evanston, Ill.: Northwestern University Press 1999.
[57] Zahavi, D.: *Subjectivity and Selfhood: Investigating the First-Person Perspective*. Cambridge, MA: MIT Press 2005.
[58] Ziman, J.: No Man Is An Island. *Journ. of Consciousness Studies* **13** (2006) 17–42.

Cesta ke kyborgovi

Jana Horáková[1], Jozef Kelemen[2]

Abstrakt. Cílem této studie je naznačit některé kulturní důsledky vzniku a rozvoje kybernetiky a informatiky. Tyto disciplíny vytvořily pojmový aparát pro myšlení o strojových a lidských uspořádáních v kategoriích analogií a daly tím strojům vlastnosti, které je učinily zcela odlišnými od strojů začátku 20. století a které je přibližují živým tvorům a myslícím lidem.

Jestli něčemu „rozumíte" pouze jedním způsobem,
pak tomu stěží rozumíte – dostanete-li se
do obtíží, nevíte, kudy kam. Reprezentujete-li
však něco několika způsoby, pak, jste-li dostatečně
frustrován, máte možnost přepínat mezi více hledisky,
až pokud nenajdete to, které se vám bude jevit jako přijatelné.

Marvin Minsky: The Emotion Machine

1 Kybernetika

Rok před vydáním své proslulé knihy přijel zakladatel kybernetiky Norbert Wiener[3] z massachusettské Cambridge do Evropy. V Británii navštívil tamější Cambridge a taky Manchester. Na obou místech se živě zajímal o britské návrhy počítačů tzv. von Neumannovy architektury. V Taddingtonu se setkal s Alanem Turingem, který se v té době rovněž zabýval konstrukcí počítačů podle architektury, kterou navrhl John von Neumann a vypracovával pro ně programovací jazyk. V Paříži na Collége de France proslovil jednu z centrálních přednášek na konferenci věnované matematice. V polovině června 1947 se vrátil na několik dní do Bostonu a pak odcestoval do Mexico City na Národní kardiologický institut, za svým kolegou a dlouholetým přítelem Arturem Rosenbluethem. Právě tam, zřejmě i pod vlivem názorů, se kterými se setkal v Evropě, napsal Wiener svou rozsahem nevelkou, významem však obrovskou knihu *Cybernetics, or Control and Communication in the Animal and the Machine*.[4]

Mohutná vlna zájmu o obecnou povahu řízení a komunikace, „manifestem" které se Wienerova kniha záhy stala, se vzedmula jako reakce na stále intenzivněji pociťovanou absenci dostatečně autentického výzkumu těchto fenoménů v podmínkách, ve kterých docházelo k nebývalému rozvoji (elektro)techniky a k významné revoluci i v oblasti komunikačních a informačních technologií. V první polovině minulého století nabízela, jak se postupně ukazovalo, tradičně pojímaná aplikovaná matematika již pouze poměrně

[1] Ústav hudební vědy, Filozofická fakulta Masarykovy university, A. Nováka 1, 602 00 Brno, ČR, E-mail: horakova@phil.muni.cz

[2] Ústav informatiky, Filozoficko-přírodovědecká fakulta Slezské university, Bezruč. nám. 13, 746 01 Opava, ČR, E-mail: kelemen@fpf.slu.cz a Vysoká škola manažmentu, Panónska cesta 17, 851 04 Bratislava, SR

[3] Informace o Wienerově životě čerpáme z [3].

[4] Předmluva k 1. vydání – které vydaly společně The Technology Press, Cambridge, Mass. a John Wiley & Sons, New York, v angličtině a Herman et Cie., Paris, ve francouzštině v roce 1948 – je datována do listopadu 1947.

omezené prostředky pro vyjadřování se o oblastech reálného světa, vynořujících se ze spousty technických, technologických a také kulturních novinek, které technický pokrok přinášel.

Bylo čím dál tím zřejmější, že z pohledu započatého technického a technologického rozvoje se *řízení* a *komunikace* stávají nesmírně důležitými aspekty chování systémů, a to bez ohledu na to, o jaké systémy se jedná. Vznik vědního oboru, který by poskytl zázemí pro sjednocující pohled na systémy rozmanité povahy a rigorózní pojmový aparát pro jejich popis, byl nezbytný. Do doby, než Wiener ustavil novou vědní disciplínu, však neexistovala žádná alespoň relativně samostatná a dobře konstituovaná vědecká disciplína, jejímž objektem by byly tyto fenomény. Doba Wienera – konec první a začátek druhé poloviny 20. století – však již nabízela všechny logické, metodologické i instrumentální předpoklady pro konstituci takové disciplíny. Vlna, která se vpádem Wienerovy knihy do oceánu vědy vzedmula jako tsunami – a dnes to již můžeme trochu i upřesnit: šlo o první vlnu zájmu takovéhoto druhu a orientace – získala své označení odvozené z Ampérova *cybernétique*.

Wiener byl matematik a otevřeně se hlásil k filozofickým názorům, které v jeho době zastával jeden z nejvýznamnějších filozofů minulého století, Bertrand Russell, k pozicím, které hlásal a hájil snad nejvýznamnější logik minulého století a dle mínění některých logiků dokonce, pokud jde o význam jeho objevů, druhý nejvýznamnější logik v dějinách po Aristotelovi (mimochodem rodák z Brna), Kurt Gödel, nebo k průkopnickému dílu jednoho ze zakladatelů matematické teorie informace, svého kolegy z MIT v massachusettské Cambridgi, Clauda Shannona. Tyto vlivy na Wienerovo myšlení dokumentují početné pasáže v jeho fundamentální knize o kybernetice.

Díky obrovskému úsilí Wienera a mnoha jeho kolegů, vesměs fundovaných specialistů z mnoha vzájemně někdy až překvapivě vzdálených oblastí vědy a techniky, dosáhla kybernetika výsledků, které mimořádně ovlivnily dnešní podobu lidské civilizace a které vydatně přispěly i k jakémusi „obrození“ významu matematiky. Otevřely nové možnosti jejím aplikacím a postavili kybernetiku do role jedné z nejvlivnějších vědných oblastí, utvářejících ve 20. století podobu civilizace západního typu.

2 Informatika

O fundamentální problémy matematiky se začátkem 30. let minulého století začal v Británii zajímat druhý později významný představitel vědy toho století, Alan M. Turing.[5] Na začátku své vědecké kariéry, v roce 1936, publikoval své představy o tom, jak matematicky precizně definovat výpočet a rozdělil tím matematicky definované funkce – např. $y = x^2$ nebo $y = sin(x)$ apod. – do dvou základních skupin – na vypočitatelné (později se vžil termín *turingovsky vypočitatelné*), tedy ty, jejichž hodnoty se dají vypočítat konečným počtem základních operací z hodnot jejich vstupních proměnných, a na *turingovsky nevypočitatelné* funkce, které tuto užitečnou vlastnost nemají. Je snad zřejmé, že po konstrukci prvních počítačů se staly ty vypočitatelné funkce prakticky velice zajímavé, protože počítače byly díky své technické konstrukci schopny Turingem předpokládané základní výpočetní operace vykonávat. Ty nevypočitatelné zůstaly pro počítače dodnes nedobytou (nebo alespoň hodně problematicky dobytnou) pevností.

[5] O jeho životě viz biografii [11].

Wiener byl o stavu techniky v časech, když psal svou *Kybernetiku*, výborně informován a je zřejmé, že svou knihu nenapsal bez vědomí existence tehdejší výpočetní techniky, a zřejmě ani bez kontaktů s osobnostmi, které stály u vzniku a počátečního rozvoje počítačů. Tato technická novinka počátku druhé poloviny 20. století se však v úvahách, které předložil ve své knize, objevuje pouze v určitém specifickém, z dnešního pohledu dosti zúženém, kontextu.

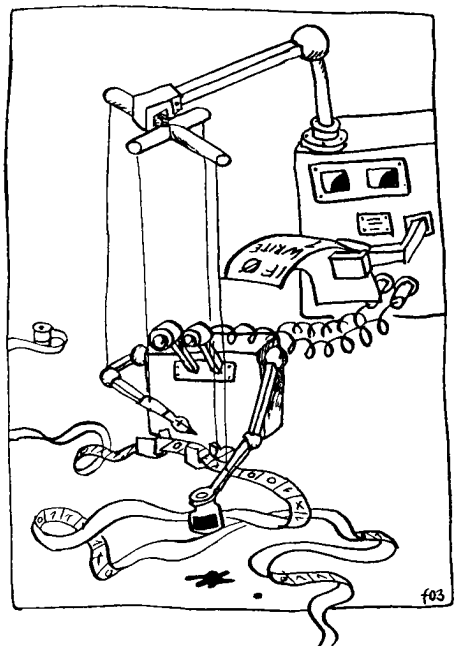
Určitá míra ignorace některých aspektů výpočetní techniky ve Wienerově uvažování o kybernetice mohla pramenit z uvědomení si rozdílnosti motivů, které ho přivedly ke kybernetice a motivů, které přivedly průkopníky výpočetní techniky ke konstrukci prvních počítačů. Jejich motivem bylo vytvořit – pod obrovským tlakem společenské objednávky plynoucí především z vývoje a konstrukce první atomové pumy – pokud možno co nejrychlejší stroj na provádění lidskými silami nezvládnutelných numerických výpočtů, na výsledky kterých čekali fyzici v novomexickém Los Alamos (a pro které byla von Neumannem navržena po něm nazvaná architektura počítačů). Wiener v 5. kapitole své *Kybernetiky* rozděluje počítací stroje v zásadě do dvou skupin – první skupinu tvoří *analogové stroje*, druhou *číslicové stroje* mající charakter (dovednějších) kalkulaček.

Wiener si dobře uvědomoval základní princip tehdejších nejmodernějších počítačů, které se vyznačovaly pamatováním nejenom dat, nýbrž i programů jejich zpracování (tedy princip tzv. *von Neumannovy architektury* počítačů). V úvahách o potřebě programovat číslicové stroje se však Wiener dostává k matematické logice a vnáší do kybernetiky problematiku návrhu logických obvodů jako elektrotechnicky zcela přirozeně uskutečnitelných konstrukcí, propojených velice přímočaře s nesmírně exaktní oblastí matematiky – s problematikou tzv. *boolovských funkcí* (tedy zhruba řečeno těch funkcí, které nabývají pouze hodnoty 0 nebo 1). V té samé kapitole uvažuje pak o tom, jak by bylo možné propojit problematiku (konceptuálních nebo technicky uskutečněných) sítí logických elementů s problematikou činnosti mozku při rozhodování nebo při učení se.

Mnohem méně se Wiener věnuje ve své stěžejní knize *teorii výpočtů* a *abstraktních výpočetních strojů*, jak byly iniciovány v již zmíněné významné práci Alana M. Turinga z roku 1936 [38], ve které Turing zavedl matematicky precizní definici pojmu *vypočitatelnosti* a *abstraktního počítače* (někdy *stroje*, resp. *Turingova stroje*).

V Turingově původním smyslu, tedy v tom z roku 1936, jsou *vypočitatelná* (angl. *computable*) taková reálná čísla, decimály kterých jsou vypočitatelné konečným počtem kroků, tedy pokud je dokáže postupně napsat stroj. Právě podoba tohoto stroje – v Turingově původním označení tzv. *a-stroje*, který pak Emil Post pojmenoval *Turingův stroj* – byla podrobně specifikována. Turing upozorňuje, že i když jsou objektem jeho zájmu čísla, svůj pojem vypočitatelnosti dokáže bez obtíží rozšířit i na vypočitatelnost matematických funkcí definovaných na reálných číslech, nebo na zjišťování logické pravdivosti predikátů, apod. Právě ve vztahu k funkcím (a predikátům) se pak pojem vypočitatelnosti stal jedním ze základních pojmů teoretické informatiky 20. století.

Wiener sice ve své knize tuto Turingovu práci zmiňuje, nevěnuje se však Turingovu stroji, nýbrž úvahám o mechanizovatelnosti myšlení, tedy problematice, která byla v tomto Turingově článku zmíněna jenom velice okrajově. Problematikou počítačů ve vztahu k myšlení se Turing zabýval podrobněji až později, v jiné své práci [39], ze které vzešel pro umělou inteligenci klíčový pojem tzv. *Turingova testu*, o kterém se ještě zmíníme.



Obrázek 1. Karikatura Turingova stroje od F. Cvrčkové z [27].

3 Postmoderní stroj

V pracích dvou výše zmíněných průkopníků nového chápání stroje se dá detekovat způsob myšlení, který formou analogií nahlíží na (funkce) lidské mysli, lidského těla a akce automatických strojů. Toto téma je reflektováno ve své moderní podobě od začátku 20. století i v oblasti umělecké tvorby.

Hru *R. U. R.*, ve které bylo poprvé použito slova *robot*, psal Karel Čapek v létě 1920 během pobytu u rodičů v Trenčianských Teplicích. Dle autorových vzpomínek mu tam jeho bratr Josef nabídl – místo Karlem původně zvažovaného, ale dle jeho vlastních slov poněkud "papírově znějícího" pojmenování umělého dělníka *labor* (slova latinského původu avšak s poměrně prvoplánovými politickými konotacemi) – pojmenování *robot*, které má slovanské kořeny, ovšem s velmi hluboko do historie zasahujícími konotacemi, slovo, které zpřítomňuje představu všech otroků, sluhů, nevolníků, dělníků, zkrátka všech těch, kteří se nechávají najímat na práci.

Metoda převýchovy („kulturní reformy“), kterou ve společné povídce obou bratrů *Systém* z roku 1908 aplikoval továrník Ripraton na své dělníky, aby je formoval směrem k ideálu – ke stroji – byla v dramatu *R.U.R.* dovedena k absurdní dokonalosti. Čapek presentuje tovární systém, ve kterém jsou dělníci nahrazeni svou umělou, a ve své účelné jednoduchosti dokonalou, náhražkou – roboty. Jeho ideální továrna je továrna bez dělníků! Autor takto ironizuje všechna dobová radikální řešení vzrůstajícího napětí mezi továrníky a dělníky, ale i obraz „ideálního dělníka“ v podobě člověka s vlastnostmi stroje. Aby odhalil nereálnost těchto vizí, používá jejich vlastní zbraně a předkládá natolik „radikální řešení“, že ve své nekompromisnosti překonává i ty nejodvážnější reformátory.

I tito umělí dělníci-stroje-roboti se však v určité fázi své emancipace svým tvůrcům a pánům vzepřou vší svou ničivou sílou a “ne-lidskostí“, které jim byly (i když

nechtěně) dány do vínku. Skoncují s lidstvem, jenž paralelně s vývojem robotů spěje svou degenerací k zániku. Vzpoura robotů zánik lidstva nezpůsobila, pouze urychlila. Poslední žijící člověk ve hře *R. U. R.*, Alquist, je však ještě svědkem toho, jak vzplane láska – emoce spojovaná s lidským rodem a s lidstvím – mezi robotem Primem a robotkou Helenou. To mu dává naději, že civilizace Robotů bude dál rozvíjet dědictví svých otců, tedy civilizace lidské.

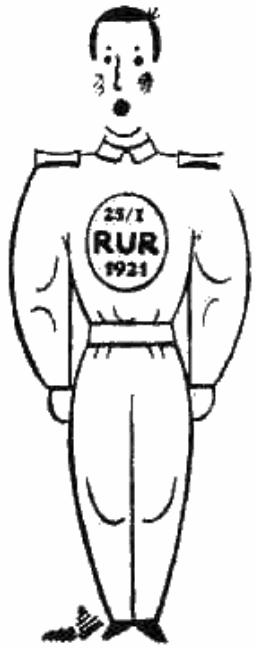
Místo a datování premiéry hry *R. U. R.* je trochu problematické. V pražském Národním divadle byla hra poprvé představena 25. ledna 1921. Hru však inscenovala již o tři týdny dříve Jednota divadelních ochotníků Klicpera na prknech Městského Klicperova divadla v Hradci Králové, a to 2. ledna 1921. Tuto zvláštní premiéru před oficiální premiérou režíroval inspektor Státních drah Bedřich Stein. První pár robotů hráli: Robota Prima F. Paclt a robotku Helenu J. Skorkovská. Obrazová dokumentace se z této premiéry bohužel nedochovala. V pražském Národním divadle se ujal režie Vojta Novák. Scénu navrhl Bedřich Feuerstein a kostýmy sám autor slova robot – Josef Čapek, pro kterého to bylo prvním setkáním s divadlem v roli výtvarníka. Role robotů Prima a Heleny hráli Eduard Kohout a Eva Vrchlická, která hrála i Helenu Gloryovou.

Josef Čapek zvolil pro roboty-dělníky velmi střízlivé a pouze mírně stylizované kostýmy v podobě modrých pracovních kombinéz, které byly jednoduše odvozeny od podoby pracovního oděvu továrního dělníka. Postava robotky písařky Sully měla obdobně neutrální ženský kostým služky. Ne-lidská podstata robotů byla vyjádřena pouze hrou herců - strnulými výrazy tváří a sekanými pohyby. Jak je uvedeno ve scénických poznámkách k dramatu: *Roboti v předehře oblečeni jako lidé. Jsou úsečni v pohybech i výslovnosti, bezvýrazných tváří, upřeného pohledu. Ve vlastní hře mají plátěné blůzy v pasu stažené řemenem a na prsou mosazné číslo*, instruoval Karel Čapek interprety jeho slavné hry [5]. Jestliže tedy Karel Čapek prezentuje v textu své hry roboty jako továrním systémem sériově produkované "výrobky", jako komoditu, jde vlastně o metaforické pojmenování motivu dělníka jako zaměnitelné jednotky v soukolí systému továrny. Postava robota však v sobě obsahuje také koncept ideálního dělníka moderní doby, který má podobu *živého a inteligentního pracovního stroje* plnícího příkazy bez výhrad a bez zaváhání, s rychlostí a s přesností automatu. Proto se také automat, stroj stává ideálem, ke kterému je dělník a nejen dělník, ale obecně pracující člověk průmyslové civilizace, přirovnáván.

Zatímco v první polovině 20. století umělecké avantgardy zobrazovaly vztah člověka a stroje jako zrcadlení dvou autonomních entit, po druhé světové válce, v souvislosti s novou podobou, kterou stroj nabývá, a kterou jeden z nás [16] nazval *postmoderním strojem*, se radikálně proměňuje i zobrazování vztahu člověka a stroje. Nové pojetí vztahu člověka a stroje vychází z pozice *kybernetického*, resp. *informatického paradigmatu* a v tomto smyslu neklade důraz na odlišnosti člověka a stroje, nýbrž chápe nové výpočetní technologie jako model funkcí lidské mysli a později jako partnery v komunikaci/interakci. Umělci, stejně jako vědci, se soustředili spíše na to, co člověka a stroj spojuje (v současnosti třeba atribut jejich systémovosti v rámci *kybernetizmu*[6] nebo připodobování člověka k počítačům v rámci *komputacionalizmu*[7]), než na to, co je odlišuje

[6] Tento pohled dobře odráží filozofická analýza, kterou poskytuje [10].
[7] O tomto proudu informuje např. (Scheutz, 2002).

(v současnosti třeba předpoklad vědomí u člověka a antropocentrická skepse vůči vědomí strojů).



Obrázek 2. Kreslená postava robota (v kostýmu jaký byl J. Čapkem navržen pro pražskou premiéru) ilustruje pasáž věnovanou hře *R. U. R.* v eseji Josefa Čapka *Homo Artefactus* (1. vyd. 1924) a je karikaturou Karla Čapka s datem pražské premiéry *R. U. R.* jako „výrobním číslem" na jeho hrudi.

Katherine Haylesová ve své knize *How We Became Posthuman* [10] presentuje postmoderní obrat v chápání vzájemného vztahu konceptů člověka a stroje právě v kontextu všeobecného akceptování informatické definice člověka, jako informace zpracovávajícího systému. Můžeme konstatovat, že věda druhé poloviny 20. století strukturálně separovala informaci od jejího obsahu a tím současně eliminovala tělo z našich úvah o podobnostech a rozdílech mezi lidskou bytostí a informačními technologiemi. Došlo k tomu v rámci Claudem Shannonem iniciované *teorie informace*. Ta nabídla matematicky fundovaný prostředek pro technicky nezbytnou kvantitativní analýzu informací, naprosto však vytrhla informace z jejich sémantického, tedy kvalitativního kontextu.

Objevuje se i pocit podobný tomu, který na začátku století zaregistrovali futuristé. V prvním *Manifestu futurizmu*[8] jeho autor Filippo Marinetti hlásal potřebu *odstranění statického, nezajímavého umění minulosti* a naléhavě zdůrazňoval potřebu nových technologií (zmiňuje konkrétně auta, letadla, lokomotivy). Složitost a síla těchto strojů je podle jeho mínění to, co by mělo diktovat estetiku a inspirovat uměleckou tvorbu budoucnosti. *„Teplo kusu železa nebo dřeva je od nynějška rozvášňují více než ženské slzy a úsměvy!" „Cítíme mechanicky a cítíme se být uděláni z oceli"*, opakuje, parafrázuje a ironizuje o něco později výkřiky futuristů autor slova „robot" i prvního vyobrazení tohoto konceptu Josef Čapek v *Homo Artefactus* [4].

[8] První manifest futurizmu byl publikován v deníku *Le Figaro* v roce 1909.

Za současného následovníka futuristů začátku 20. století můžeme považovat proslulého australského performera vystupujícího pod jménem Stelarc, který ve svém manifestu kyborga napsal: *Nemá již dále smysl vidět tělo jako sídlo naší psychické a společenské existence. Je spíše strukturou, kterou je třeba monitorovat a modifikovat – nechápejme tělo jako subjekt, nýbrž jako objekt, a nechápejme ho jako objekt, po kterém toužíme, nýbrž jako objekt, který vytváříme*.[9] Na rozdíl od umění futuristů, kteří byli fascinováni mechanickým a byli připraveni hledat k němu cestu z domnělé beznaděje lidsky organického, Stelarc jako kdyby hledal místo ´nového člověka´ kdesi uprostřed metaforického šera mezi přirozeným (biologickým) a umělým (mechanickým). *V konfrontaci s obrazem své sešlosti je tělo traumatizováno a snaží se odštěpit od říše subjektivity zvažujíce nevyhnutelnost svého přehodnocení a možnost nového navržení své vlastní struktury*, hlásá Stelarc. A souhlasil by s ním zřejmě i vědec: *Možná nadešel čas zvažovat přetvoření samých sebe do forem, které poskytují větší možnost individuálního růstu. Mohli bychom uvažovat o možnosti převést se do podoby strojů*, píše Marvin Minsky [30] a zcela konkrétní pokusy takovéhoto zvláštního sebezdokonalování, posouvání se od krajnosti s nápisem *člověk* směrem ke středu zmíněného intervalu, propojením vlastního těla (a do jisté míry i mysli) se strojem nám čas od času nabízí i jiný specialista v oboru informatiky a robotiky, Kevin Warwick z univerzity v britském Readingu[10].

4 Médium a interaktivita

V souladu s jeho postmoderním chápáním se stroj stal v dějinách kultury postupně nejenom metaforickou pomůckou pro literární, divadelní[11] nebo kinematografickou tematizaci určitých společenských jevů, problémů, názorů a estetických koncepcí. Stal se i *médiem* pro rozsáhlou oblast umění, která se dnes nazývá mediálním uměním a jehož počátky jsou všeobecně kladeny do 60. let 20. století.

V posledních několika dekádách 20. století byly v oblasti výzkumu a vývoje digitálních médií realizovány některé klíčové vize moderního umění, a to dokonce přibližně ve stejné době, kdy byly tyto vize formulovány umělci. Paralelismus či synchronizace tohoto vývoje zasahuje nejen obsah, tedy koncepty, ale i podoby, v jakých jsou tyto myšlenky formulovány. Ve 20. století umělci charakteristicky prezentovali své myšlenky buď psaním manifestů nebo vytvářením aktuálních uměleckých děl, zřetelně směřujících od objektu ke konceptu (mluvíme o dematerializaci umělecké tvorby). V případě počítačových expertů máme k dispozici buď teoretické články, které rozvíjejí návrhy konkrétních softwarů nebo/a hardwarů (což jsou jakési analogie k manifestům), nebo více popisné články o již vytvořených prototypech nebo již fungujících systémech (což můžeme chápat jako pandán k aktuální umělecké tvorbě).

V obou sférách poznání můžeme mluvit o zaměření na vlastní proces vnímání uměleckého díla (interaktivitu), se kterým souvisí tendence k dematerializaci uměleckého díla, a o zrušení hierarchie vztahu autor vs. vnímatel ve prospěch otevřenosti uměleckého

[9] www.stelarc.va.com.au

[10] Poutavý obraz o takovýchto svých experimentech nabízí ve své knize [42].

[11] V čase když píšeme tuto studii, si kulturní svět připomíná 85. výročí prvního použití slova *robot* (premiérově byla jeho hra *R. U. R.* uvedena v pražském Národním divadle 24. ledna 1921). My jsme tomuto výročí věnovali článek [13].

zážitku (více nebo méně očekávaným) intervencím vnímatelů a jeho formování v reálném čase probíhající interakce. Co však odlišuje uměleckou tvorbu v prostředí *interaktivních médií* od ostatní tvorby (viz aktivity hnutí *Fluxus*, happening, performance art, konceptuální umění) je skutečnost, že výše popsané postupy a strategie vytváření uměleckého prožitku nejsou výrazem postoje umělce k umělecké tvorbě a jeho chápání umění obecně, nýbrž jsou zakódovány v samotných principech tvořících bázi interaktivního média. Dieter Daniels v této souvislosti upozorňuje, že zatímco v 60. a 70. letech 20. století můžeme mluvit o uměleckých aktivitách, které ovlivňují produkci médií, přibližně od 90. let se situace obrací a tvorba umělců je primárně inspirována technickými možnostmi, které jim nová média poskytují [6].

Umění interaktivních či nových médií je samotnou podstatou protikladné k tradiční představě o uměleckém díle, jako jedinečném a neopakovatelném objektu, a o umělci, jako geniálním tvůrci. Umělecké dílo vytvořené v prostředí interaktivních médií má potencionálně nekonečnou řadu podob, které nabývá skrze a v procesu interaktivity; proti jedinému autorovi staví kolektivní, kolaborativní autorství; proti fetišizaci uměleckého originálu staví nekonečnou řadu reprodukcí či variací, z nichž žádná nemůže být považována za definitivní nebo jedinou „pravou“ verzi interaktivního díla. Tyto principy, které jsou analogické anti-uměleckému hnutí 60. let, nejsou pouhou deklarací určitých preferencí umělců, nýbrž vycházejí ze samotných principů tvořících základ *interaktivního média* jak je analyzuje americký umělec a teoretik nových médií Lev Manovich [26].

Jednoduchým a výjimečně podmanivým příkladem interaktivního média byl *Kinoautomat* Radúze Činčery, žnoucí zasloužený úspěch na Expo ’67 v kanadském Montrealu v roce 1967. Šlo o filmový příběh s nádechem černého humoru postavený na sérii flashbacků muže, který se činí zodpovědným za požár svého bytu. Po každé ucelenější sekvenci filmu se před plátnem objevila osoba a požádala diváky, aby hlasovali o tom, kterou ze dvou nabízených možností má příběh pokračovat. Hlasovalo se tlačítky spojenými se zelenými a červenými světly obklopujícími promítací plátno. Výsledek hlasování byl takto viditelný a většina rozhodla o pokračování děje, který uskutečňoval kinooperatér, obsluhující dvě synchronizované promítací aparatury. Děj se však ve skutečnosti nevětvil jako binární strom. Jakkoli se totiž diváci rozhodli, k následujícímu hlasování došlo v situaci, která byla stejná pro obě dříve volitelné alternativní pokračování. Technicky bylo celé představení realizováno bez použití počítače. Ty byly v časech *Kinoautomatu* ještě pro takovéto použití příliš veliké a (zejména v československých podmínkách) příliš nákladné. Nic to však nemění na skutečnosti, že někteří teoretici umění interaktivních médií, jmenovitě např. Michael Naimark [32], pokládají *Kinoautomat* za jedno z doposud nejzdařilejších děl v této oblasti umění.

Počítač se stal poprvé součástí interaktivního uměleckého projektu pravděpodobně v díle Myrona Kruegera. Ten začal v roce 1969 se svými spolupracovníky na Wisconsinské univerzitě vytvářet interaktivní dílo *Glowflow*. Bylo to prostředí místnosti, ve které snímaly pohyb návštěvníků senzory vbudované do podlahy a citlivé na tlak. V rozích místnosti byly umístěny reproduktory a po stěnách trubice naplněné suspenzemi různých barev. Pohyb návštěvníků byl vyhodnocován počítačem a výsledky vyhodnocení rozhodovaly o tom, jaké zvuky budou vydávat reproduktory a které trubice se rozsvítí[12].

[12] Více o zkušenostech M. Kruegera s projektem *Glowflow* viz v [20].

Pokoušíme-li se postihnout, co dává interaktivním uměleckým projektům nebo dílům jejich zvláštní působivost, čím nás oslovují, z čeho čerpají své zvláštní kouzlo, možná mnozí budeme souhlasit s tím, že je to právě jejich připravenost *interagovat*. Je to jejich specifická „neukončenost“ nebo „neuzavřenost“, která je příslibem, že se v rámci interakce s dílem může něco nepředvídatelně „přihodit“, že každého interagujícího adresáta můžou zastihnout individuální a možná již nikdy jindy nikomu jinému se nevynořící zážitky, který svou jedinečností může v konkrétním pozorovateli vyvolat *zcela* individuální estetické, etické, emotivní, nebo kognitivní prožitky.

Jako příklad uvedeme „starý ale dobrý“ interaktivní video-projekt Bena Pipera: Kamera snímá návštěvníky expozice, kteří se objevují na obrazovce na bílém pozadí. Náhodně vybírané sekvence jsou pak viditelné spolu s právě aktuálními. Čím delší dobu je sekvence uchovávána v paměti, tím je „slabší“, až se stane „neviditelnou“. Celá instalace funguje tedy jako jakási „paměť“, která pomalu ale jistě „zapomíná“. Jak napsal Norman E. Spear: *...obsah paměti je funkcí rychlosti zapomínání*[13]. Současně se před námi objevuje jakýsi prostor-scéna, zprostředkující představu fungování lidské mysli – snění, odpoutaného od zákonitostí reality, ve kterém se mohou propojit záznamy prožitků a obrazy osob, které by se v kauzální příčinnosti časoprostoru, ve které žijeme, nikdy nemohly setkat.

Jiným příkladem zdařilé interaktivní instalace je projekt *A-Volve* Christy Sommererové a Laurenta Mignonneaua s vodní hladinou, pod kterou je instalována obrazovka počítače s živým rybám podobnými virtuálními kreaturami a technické zařízení na zjištění poloh předmětů ponořených do vody. Kreatury mají individuální chování, v jehož rámci reagují třeba na pohyby rukou ponořených do vody nad obrazovkou. Vzniká tak dokonalá iluze interakce kreatur s pozorovatelem. Systém umožňuje vytvářet si kreatury dle vlastních figurativních a behaviorálních představ a interagovat pak i s vlastními výtvory[14]. Tekutost rozhraní mezi evokovaným vodním světem a návštěvníky poukazuje k tekuté povaze *„plovoucího uměleckého díla“*[15] (v protikladu k otevřenému dílu ve smyslu Umberta Eca) v prostředí nových médií, *plovoucího díla* ve smyslu jeho potencionální neukončenosti a neustálé otevřenosti vůči možným intervencím do světa za obrazovkou jako hlavního principu interaktivní umělecké tvorby.

Interaktivita či přívlastek *interaktivní* nemá v oblasti umění svou jasnou a obecně přijímanou definici.[16] Původně se interaktivita objevuje v souvislosti s novými médii jako jejich charakteristická vlastnost, časem se však stává spíše prostředkem tematizace obecného problému poznání lidského vnímání a užívaných způsobů komunikace. Lister a jeho kolegové chápou například interaktivitu jako *...schopnost* [která je] *poskytnutá uživateli, který může zasahovat do výpočetních procesů a vidět výsledky těchto zásahů v reálném čase*. Pojmu je však používáno, jak připomínají, také *...v teorii komunikace k popisu lidské komunikace založené na dialogu a výměně* [24]. Interaktivita odkazuje tedy nejobecněji ke dvěma oblastem – *k strukturám* a *procesům* realizovaným digitálními médii, které můžeme poměrně snadno popsat prostředky, které nabízejí příslušné oblasti

[13] Norman E. Spear v předslovu k souboru esejů [41].
[14] Interpretaci a zhodnocení projektu *A-Volve* nabízí [43].
[15] Pojem Söke Dinkla *„plovoucí umělecké dílo“* uvádí [34].
[16] Viz též [12].

informační techniky a technologie[17]. Odkazuje však i k teorii lidské interakce a komunikace, jejíž komplexnost nás předurčuje k vyjadřování se v podobě metafor a k vytváření modelů našich představ o možných principech, které ji formují. Ne v poslední řadě nás pak odkazuje i k problematice interakce mezi člověkem a strojem. Jak míní Erki Huhtamo, *...interaktivita je stále primárně mentální a intelektuální a pouze druhotně technologická aktivita* [14].

Skutečně platí, že bez toho, *kdo* interaguje a *kdo* má schopnost interpretovat výsledek interakce, je samotný fenomén součinnosti z hlediska (umělecké) komunikace bezpředmětný. Zajímavou se však součinnost v tomto kontextu stává, zaměníme-li zájmeno „*kdo*“ za *obecnější* „*co*“. Ta malá změna nás přivede k problému umění entit vybavených určitou mírou inteligence i emotivity a to bez ohledu na to, jsou-li tyto entity lidské bytosti nebo ne. Otevírá se tak zajímavá možnost poměrně dobrodružné úvahy nejenom o robotickém umění, nýbrž i o *umění robotů či strojů obecně....*

5 Kognice a emoce

Fakt, že se ze stroje stalo jednak mocné médium mezilidské komunikace (člověk – síť – člověk), ale i partner v interakci člověk-stroj, představuje výzvu nejenom pro ty, kdo stroje resp. možnosti jejich interakce s lidmi využívají jako prostředek pro zkvalitnění některých lidských činností, schopností nebo prožitků, nýbrž i pro ty, kteří zkoumají a uskutečňují možnosti a způsoby zdokonalování strojů a v rámci postmoderního chápání stroje (o kterém jsme se zmínili výše), i pro ty, kterých zájem se soustřeďuje na hlubší a komplexnější poznání člověka. Podstatným se tak i pro nás stane pohled těch, kteří hledají způsob, jak zkoumat interakce stroje a člověka tak, abychom se pokud možno co nejméně soustřeďovali na odlišnosti mezi člověkem a strojem a co nejvíce zdůrazňovali, co je v rámci jejich interakcí spojuje nebo může spojovat.

V kontextu *Turingova stroje*, o kterém jsme se zmínili na začátku, může být formulován způsob zjišťování možnosti programovat počítače tak, aby jejich činnost byla „navenek“ (nebo, jak se někdy říká ve filozofii mysli, z „pohledu třetí osoby“, angl. *third person view*) nerozeznatelná od inteligentní činnosti člověka. Tomu způsobu se v oblasti umělé inteligence říká *Turingův test*. V jeho pozadí stojí otázka, je-li inteligence projevem výpočtů soustavy funkcí vypočitatelných na *Turingově* (nebo na libovolném s ním ekvivalentním abstraktním) *stroji*. Je tedy inteligence v zásadě matematicky definovatelnou a turingovsky vypočitatelnou funkcí? Je inteligence něco ve své nejhlubší matematické a výpočetní podstatě stejného jako funkce $y = x^2$ a výpočet jejích hodnot pro různé hodnoty proměnné x, jenom mnohem komplikovanější? Je inteligence skutečně proces, do kterého *na začátku* vstoupí data (coby hodnoty proměnných x, třeba jako kód popisující určitý problém) a *na konci* je produkován výstupní kód řešení toho problému (jako hodnota y vypočítaná danou funkcí – přesněji *Turingovým strojem*, který jí odpovídá – z příslušných vstupů)?

V průběhu druhé poloviny 20. století se důležitým impulsem zdokonalování strojů stala snaha, aby obstály v *Turingově testu*, tedy přibližování jejich schopností schopnostem a dovednostem člověka. Tato snaha nabyla svou konstituovanou podobu zejména

[17] Lev Manovich, teoretik vycházející z počítačové vědy, popisuje kromě „uzavřené“ a „otevřené“ interaktivity několik dalších konceptů jakými jsou např. tzv. menu-based interactivity, salability, image-interface, image-instrument apod. [26].

v oblasti *umělé inteligence* a *pokročilé* (*kognitivní*) *robotiky*. Problém, jak odstranit propast mezi strojem a člověkem, však zřejmě nemůže být řešen pouze zdokonalováním strojů, nýbrž musí být veden z obou stran. To znamená, že i vlastnosti člověka musí být podrobeny novému druhu zkoumání. Takovéto zkoumání musí být vedeno z pozice, která umožní vyjadřovat se o lidských schopnostech a dovednostech a o dovednostech a schopnostech strojů ve stejném pojmovém rámci. Znovu se dostáváme k pozici, kterou vhodně vymezuje oblast bádání nesoucí označení *kognitivní věda*[18].

Doposud trvají diskuse v oblasti *filozofie mysli* o tom, co na poli autonomní racionality a kognice dokáží lidé a co (ne)dokáží stroje. Jsme díky naším myslím něčím výjimečným a ojedinělým v přírodě, nebo jsme – to je druhý extrém – v podstatě stroje? Někteří – např. Rodney Brooks[19], jedna z vedoucích odborných autorit světové pokročilé robotiky – se přiklánějí k názoru, že jsme v podstatě stroje. Můžeme mu dát za pravdu, ale nemusíme. Nepoznáme přesný obsah pojmu *mysl*, a kupodivu ani pojmu *stroj*. Co si pod ním představují třeba Roger Penrose[20], David Chalmers[21], Daniel Dennett[22] nebo John Searle[23], dnes snad nejvlivnější účastníci diskusí o strojové inteligenci a vědomí strojů? A co si o názorech těchto filozofů myslí konstruktéři „inteligentních strojů“?

Lidé jako Rodney Brooks nebo např. Kevin Warwick a samozřejmě mnoho dalších v mnoha laboratořích umělé inteligence, kognitivní vědy a pokročilé robotiky po celém civilizovaném světě, sestrojují stále dokonalejší, šikovnější, svým prostředím ve svých reakcích přiměřenější, tedy inteligentnější stroje. Konkrétní, hmatatelné stroje, jejichž schopnosti můžeme ověřovat, dáme-li si trochu námahy. Na straně filozofů mysli jsou zase zajímavé hypotézy. Hypotézy, které v převážné většině případů zahalují mysl do spekulativní komplikovanosti kvantových dějů (Penrose), transformující mysl na hledání čehosi fundamentálně nového v univerzu (Chalmers), dogmaticky tvrdící, že nic kromě člověka mysl mít nemůže (Searle). Dennettův názor – velice nepřesně ho vymezím jako přesvědčení, že existují různé úrovně mysli, mysl velice vysoké úrovně má člověk, nižší někteří jiní živočichové a neměli bychom odmítat ani předpoklad, že někde velice nízko je úroveň mysli strojů – je trochu polemičtější. A tak je asi oprávněné se ptát, kdo z inženýrů a filozofů hledá pravdu o lidech a strojích a kdo pouze způsob co nejúčinnější obrany převládající ideologie doby vypjatého antropocentrizmu?

Ať je hledajícím kdokoliv a nechť hledá cokoliv, jeho situace není dnes, na začátku 21. století, snadná! Ocitli jsme se na hranici mezi přírodním a umělým, živým a neživým, vědomým a tím, co vědomí nemá. A otázka rozdílu už není pouze gnozeologická. Pomalu, ale jistě se stává problémem ontologickým. Použijíc pojmenování antropologa Victora Turnera [40], je to, kde jsme se v čase ocitli, *liminálním* okamžikem – okamžikem přechodu, okamžikem, kdy vyvstávají nové kulturní symboly a významy, momentem intelektuálního napětí, extrémních emocionálních reakcí, ale i nových možností a příležitostí konat.

Všechny tyto diskuse mají svůj význam. Postupně se v jejich rámci zpřesňovaly a dále zpřesňují významy několika klíčových pojmů, se kterými se často operovalo

[18] Úvod do takto chápané kognitivní vědy nabízí vysokoškolská učebnice [37].
[19] Viz jeho knihu *Flesh and Machines* [2].
[20] Svoje názory představil např. v knize *Makrosvět, mikrosvět a lidská mysl* [33].
[21] Vlastní koncepci mysli a vědomí předložil v monografické podobě v knize *The Conscious Mind* [15].
[22] Určité shrnutí jeho názorů na mysl najde čtenář ve slovenštině v knize *Typy myslí* [7].
[23] Jeho názory zprostředkovává např. kniha *Mysl, mozek a věda* [35].

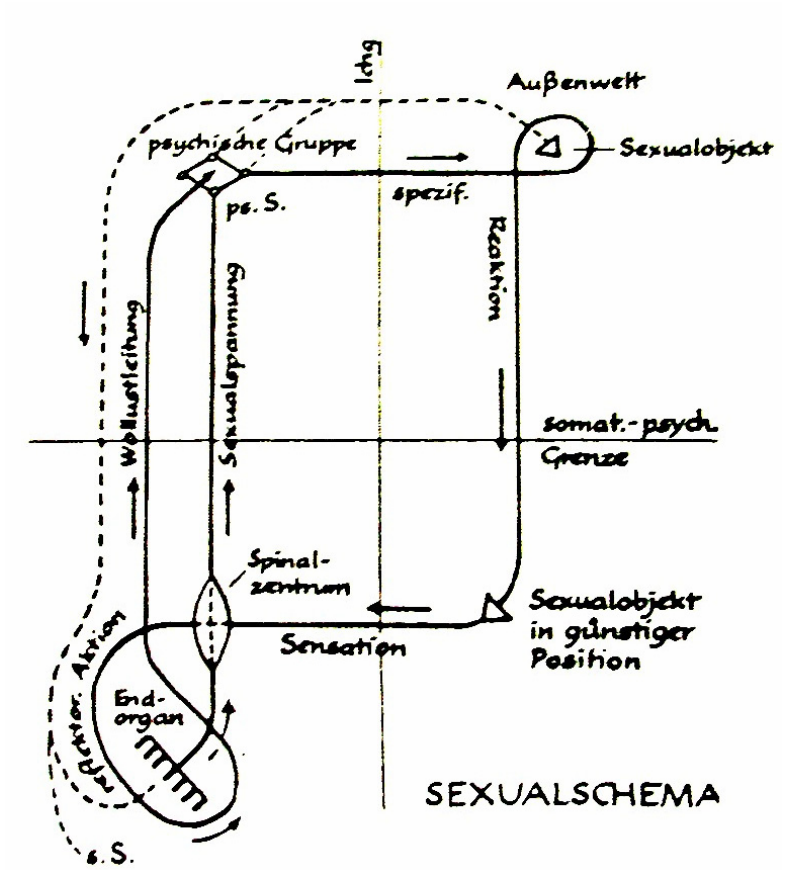
a operuje a kterých význam je nejasný a dokonce časem, při zachování vší nejasnosti, ještě i proměnlivý. Především se zpřesňoval a zpřesňuje význam pojmu *stroj*, ale i pojmů *inteligence*, *racionalita*, *poznatek* a jeho *reprezentace*, *uvažování*, *učení* (*se*) a v současnosti dokonce i pojmů jako *socializace* a *emotivita*. Mnohé z toho, co se v umělé inteligenci a v oblasti pokročilé robotiky podařilo částečně přetransformovat do podoby technicky fungujících a aplikovatelných řešení, mělo svůj původ či alespoň inspiraci ve výsledcích výzkumu nejen matematiků, logiků a informatiků, nýbrž i kognitivních a inženýrských psychologů, lingvistů a z nezanedbatelné části i úvah filozofů.

První článek o mysli, chápané jako *společenství* specializovaných *interagujících* výpočetních *agentů*, publikoval M. Minsky v podobě článku [28] a pak uceleněji ve zvláštně koncipované knize [29] sestávající z 270 jednostránkových, srozumitelných a vzájemně na sebe odkazujících esejů. V knize jsou přítomny i první náznaky problematiky emotivity lidí a strojů, kterým Minsky věnoval svou monografii *The Emotion Machine* [31]. Interakce jsou tedy možná velmi aktivním prvkem fungování přirozených i umělých myslí, který je třeba brát v potaz, pokoušíme-li se hledat jakési (výpočetně formulovatelné) vysvětlení nastávání emotivních stavů a jejich souvislost s kognitivní složkou myslí.

Důležitost interakcí ve spojitosti s mechanizmy vzniku emocí bral již Sigmund Freud velice vážně v potaz, když se pokoušel formulovat základy zamýšlené „vědecké psychologie“, pod čím rozuměl, zjednodušeně řečeno, psychologii budovanou po vzoru přírodních věd[24]. Čteme-li si dnes jeho náčrt mechanizmu vzniku a vzrůstání pohlavní touhy, jenom stěží se ti z nás, kteří něco znají o agentech, ubrání dojmu, že vše, co nám chce Freud sdělit jeho jazykem, by bylo mnohem jednodušeji sdělitelné v jazyku teorie agentů a multiagentových systémů [21]: registrace (percepce) objektu sexuálního zájmu, na základě výsledku percepce aktivace sexuálních reprezentací ve vědomí, stimulace příslušných mozkových center, následkem jejich činnosti produkce substancí majících za následek somatické změny podporující sexuální akt, těmito změnami vyvolaná opětovná stimulace aktivity sexuálních reprezentací atd. až po event. uskutečnění aktu (Freud nazval tento mechanizmus „pumpou rozkoše“). To vše by asi šlo, možná poměrně jednoduše, vyjádřit i v informatickém pojmovém rámci[25] jako interakce jednoduchých (možná reaktivních?) agentů.

Minsky charakterizuje svůj přístup k emocím, který je podrobně prezentovaný v *The Emotion Machine* následovně: *Každý z našich hlavních „emočních stavů“ resultuje ze zapnutí některých zdrojů*[26] *a z vypnutí jiných – a tedy ze změn způsobů, jakými se naše mozky chovají* (s. 4). Základní zdroje mohou být, dle Minského mínění, např. *hněv*, *strach*, *hlad* a *žízeň*, mající v komplikovanějších živých organizmech určité své fyziologické koreláty a plnící důležité funkce: Tak třeba *hněv* a *strach* vedou k *sebeochraně* organizmů, zatímco *hlad* a *žízeň* k *vyživování* organizmů. To jsou však již základní prvky racionálního chování organizmů!

[24] Učinil tak v roce 1895 ve studii *Entwurf einer Psychologie, Psychologie für Neurologen*, za jeho života nepublikované; publikované viz [8].
[25] Např v takovém, jaké nabízí [25].
[26] Minsky se v *The Emotion Machine* programově vyhýbá použití pojmu *agent*, ktorý považuje již za příliš mnohovýznamový, a nahrazuje jej pojmem *zdroj* (angl. *resource*). Při našem výkladu v této stati však Minskym zdůrazňovaný rozdíl nehraje roli.



Obrázek 3. Freudovo schéma „pumpy rozkoše"; překresleno z jeho rukopisu G. Fichtnerem a publikované v [36].

Souvisejí-li však inteligence a emotivita s obecně chápanými agenty tak úzce, jak tomu nasvědčují současné zkušenosti v oblastech kolektivní robotiky nebo emotivity strojů, pak nás to může utvrdit v odůvodněnosti kratičké úvahy, kterou jsme si dovolili komentovat v závěru 4. části této studie Huhtamovo vymezení pojmu *interaktivita*.

6 Kyborg

Stroj jsme doposud předvedli v několika podobách, ve kterých posloužil rozvoji naší představivosti, našeho materiálního světa, naší (sebe)reflexi v uměleckých dílech rozmanitého druhu, i našemu chápání procesů určitých druhů, především pak výpočetních procesů. Za předpokladu určitého formálně precizního chápání stroje – třeba v podobě tzv. *Turingova stroje* – existují, jak jsme to naznačili, určité hranice toho, co je takovýmto strojem, tedy v tzv. *turingovském smyslu*, vypočitatelné.

Koncem 20. století se však tradiční představa o adekvátnosti univerzálního výpočetního zařízení – tedy *Turignova stroje* – potřebám precizního, matematicky formalizovaného vyjadřování se o strojích v podobě, v jaké jsme se s nimi stále častěji potkávali, začala postupně problematizovat. Jedním z důvodů bylo zjištění, že ne vždy je nejvýstižnější představou o fungování reálných počítačů takové jejich chápání, v rámci kterého figurují jako zařízení, *transformující vstupní hodnoty* proměnných matematicky

definovaných funkcí *konečným počtem elementárních operací* (tzv. *kroků*) dle přesně daného návodu takovéto transformace – *algoritmu* – na příslušné *výstupní* funkční *hodnoty* těchto funkcí. Reálné výpočetní systémy se postupně stávaly mnohem otevřenějšími vnějším vlivům, různým *interakcím*, což původní představu o počítání zčásti zpochybňovalo. Pochybnosti vznikaly i pokud jde o přiměřenost konceptuálního modelu reálných počítačů v podobě *Turingova stroje* jejich fyzikálně (technicky) uskutečněným podobám. Objevily se názory, že právě *ztělesněnost* (angl. *embodiment*) reálných počítačů může podstatným způsobem měnit jejich schopnosti ve srovnání s těmi, na které usuzujeme ze schopností *Turingova stroje* a s ním ekvivalentních formálních modelů.

Dnes se postupně objevují formální modely výpočtů, které nabízejí rozšíření jejich tradičního pojímání v přibližně podobném smyslu, v jakém se fyzice podařilo „rozšířit" tradiční newtonovskou mechaniku do podoby relativistické a kvantové mechaniky. Takovéto pokusy o „rozšíření" pojmu *vypočitatelnosti* se dějí v různých směrech.[27] V zásadě všechny stojí na tom, že do modelu výpočtu v nějaké podobě zahrnou i cosi turingovsky nevypočitatelné, ale teoreticky dobře zkoumatelné. Třeba často představu zařízení, které je schopno v některých momentech výpočtu zodpovědět s pomocí tzv. *orákula* určitý typ otázek, nebo *ontologickou* (tedy generátory náhodných čísel nenapodobitelnou) *náhodnost* chování se agenta[28], jeho prostředí, nebo nepředvídatelnou *masivní interakci* agentů s jejich prostředím apod.

Problematika, která má pro další rozvoj kulturních aktivit význam minimálně srovnatelný s významem kybernetiky, nových médií a informační techniky, a kterou zastřešuje název *umělý život* (angl. *artificial life*, AL), je širší než to, co spadá do oblastí zmíněných v předešlém odstavci. Oblast umělého života ovládá snaha o vysvětlení projevů života jaké pozorujeme v přírodě (počínaje zkoumáním původu biochemického metabolismu, až po evoluci strategií chování biologických společenstev) a také vlastností života v podobách, „v jakých by mohl existovat". Názvem *umělý život* bývá tedy označována i oblast, studující systémy vytvořené člověkem, které vykazují chování charakteristické pro živé systémy. Takto vymezil v roce 1989 oblast Christopher G. Langton [22], průkopník koordinovaného a dnes již částečně i institucionalizovaného výzkumu této atraktivní i provokativní oblasti vědy.

V této souvislosti se nabízí otázka: Kdybychom nabyli přesvědčení, že třeba v datovém prostředí globální počítačové sítě, něco začalo žít? Jaký počítačový virus, nebo cokoli jiné, bychom byli ochotni prohlásit za živé? V zásadě je těžké posuzovat relevanci modelu, kdy není zcela přesně vymezen objekt modelování. Na základě dosavadních zkušeností se můžeme pouze domnívat, že tvorba modelů ne zcela přesně vymezených objektů modelování může přispět k hlubšímu pochopení modelovaného. Nové chápání „živosti", nespojující jeho podstatu pouze s biochemií, nýbrž třeba i s interakcemi „živých" entit s jejich prostředím, nebo s jejich vzájemnými interakcemi a jejich komunikací a kooperací, nebo jenom s pouhou kohabitací ve sdíleném prostředí, pak nabízí řadu možností kreovat systémy, které budou ve smyslu takovéhoto chápání *živé*.

---

[27] Přehled takových přístupů poskytuje např. časopis *Theoretical Computer Science* **317** (2004) No. 1- 3, pp. 1-267, sestavený M. Burginem a A. Klingerem.

[28] Podrobnosti k tomu viz např. v [19].

Asi v 90. letech minulého století začali mediální umělci vytvářet díla inspirovaná předpoklady a postupy oboru *umělý život*.[29] Tato díla bývají zahrnována pod různé zastřešující názvy (A-life art, genetic art, generic art, evolution art a podobně). Oblast výzkumu umělého života nabídla umělcům něco doposud nemožného – *metakreativitu*. Umělci vytvářejí systémy napodobující vlastnosti živých systémů, systémy, které „mutují", „vyvíjejí se" a projevují se jakoby byly „živé". Předpoklady oboru umělý život rezonují s principy otevřeného, interaktivního, nebo „tekutého" uměleckého díla jak byly rozvíjeny v umělecké tvorbě 20. století a dovolují je dovést do důsledků. Umělci již netvoří umělecká díla v tradičním smyslu slova, ale spíše generativní a kreativní procesy. Kreativita se tak stává metakreativitou.

Whitelaw ve své knize představuje umění umělého života v jeho čtyřech charakteristických postupech a tendencích zahrnujících: využití umělé evoluce ke generování obrazů a forem v procesu nahrazujícím umělcovu kreativitu; komplexní formy interaktivních systémů, které vtáhnou diváky do umělých ekosystémů; umělecké aktivity, vycházející z předpokladů Rodneyho Brookse („bottom-up" robotics) a vytvářející vtělené autonomní společenství agentů. Koncept „abstraktního stroje" v oblasti umělého života již nezdůrazňuje biologické analogie, ale využívá techniky jaké nabízejí například buněčné automaty, aby prozkoumaly strukturu, formu a morfogenezi. V této souvislosti se opět vrací již naznačená představa umění, tvořeného ne prostřednictvím strojů, ale samotnými inteligentními či emotivními stroji.

Věda druhé poloviny 20. století významně eliminovala tělo z kontextu informatických a kybernetických úvah. Došlo k tomu tak, že – třeba v rámci Claudem Shannonem iniciované *teorie informace* – strukturálně separovala informaci od jejího obsahu. Ta nabídla matematicky fundovaný prostředek pro technicky nezbytnou kvantitativní analýzu informací, naprosto však vytrhla informace z jejich sémantického, tedy kvalitativního kontextu. Pozoruhodnou filozofickou reakcí na tento stav je jakési postkybernetické oživení fenomenologie Henriho Bergsona, projevující se např. (a snad doposud nejzřetelněji) v díle Marka Hansena [9]. Jeho přístup k tělu je veden přesvědčením, že specificky lidsky pociťovaný prostor a čas vyvstávají v lidském vědomí přes bytí lidského těla v jeho prostředí. Toto přesvědčení má pak významný vliv na jeho hodnocení interaktivního umění jako celku i jednotlivých interaktivních uměleckých projektů. V zásadě celá Hansenova monografie, ve které představuje svůj nový filozofický pohled na tělo, je postavena na analýze řady interaktivních projektů spadajících do oblasti nových médií (např. Jeffrey Shaw, Tamás Velizcky, Teresa Wenberg, Craig Kalpankjian, Douglas Gordon a jiní) a na kritickém hodnocení jiných alternativních pozic, dominujících současné vědě o nových médiích, a filozofii současného umění (např. Lev Manovich, Roy Ascott, William Michell, Friedrich Kittler, Peter Weibel a jiní). Můžeme obecně říci, že ani u Hansena nejde o návrat k tělu jako k objektu, ale spíše o zdůraznění funkcí a vlastností těla. Těla (a smyslů) nejen jako rozhraní mezi myslí (jejíž analogií je počítač) a vnějším prostředím, ale o rozšíření analogií mezi lidskou myslí a výpočty a o funkční analogie mezi lidským tělem a počítači (hardwarem), ze kterých částečně vzešel i koncept „abstraktního stroje".

[29] Podrobněji o jejich úsilí a výsledcích pojednává monografie [43]. Poměrně obsáhlou recenzí této knihy je [18].

V obecné rovině může oblast umělého života s jejím klíčovým pojmem „emergence“ otevřít prostor pro demystifikaci přirozeného života. Po této demystifikaci pak můžeme očekávat určité přeuspořádání dnes tradičně za neměnné považovaných etických a axiologických hodnot současné civilizace západního typu[30]. V rámci ní pravděpodobně získají na důležitosti „nebiologické“ atributy života, lidského způsobu života a možná i jiných způsobů životů.

Setkání a nové propojení vědy a umění tak – zdá se – plní svůj odvěký úkol: Nemění pouze umění, nýbrž i všechny, jimž je souzeno se s uměním potkávat. Nové je snad to, že člověk se dnes neproměňuje pouze na kulturní úrovni, nýbrž, díky technice, začíná měnit sám sebe i na úrovni „somatické“, na úrovni utváření svého těla. Takto změněný člověk si tvoří svou vědu, svou techniku i své umění. Pro sebe. Nebo i pro kyborgy?

Literatura

[1] Aukstakalnis, S., Blatner, D.: *Reálně o virtuální realitě*. Brno: Jota 1994.
[2] Brooks, R.: *Flesh and Machines.* New York: Pantheon Books 2002.
[3] Conway, F., Siegelman, J.: *Dark Hero of the Information Age – In Search of Norbert Wiener, the Father of Cybernetics*. New York: Basic Books 2005.
[4] Čapek, J.: Homo artefactus. In: *Rodné krajiny* (V. Binar, sest.). Praha: Mladá fronta 1985, s. 107-157.
[5] Čapek, K.: *Dramata*. Praha: Československý spisovatel 1992.
[6] Daniels, D.: Media-Art/Art-Media. In: *Medien Kunst Netz/Medienkunst im Überblick* (R. Frieling, D. Daniels, eds.). Vienna: Springer 2004, pp. 26 - 33.
[7] Dennett, D.: *Typy myslí.* Bratislava: Archa 1997.
[8] Freud, S.: Project for a scientific psychology. In: *The Origins of Psycho-Analysis* (M. Bonaparte, A. Freund, E. Kirs, eds.). New York: Bacis Books 1954, pp. 337-445.
[9] Hansen, M. B. N.: *New Philosophy for New Media.* Cambridge, Mass.: The MIT Press 2006.
[10] Hayles, K. H.: *How we Became Posthuman.* Chicago, Ill.: University of Chicago Press 1999.
[11] Hodges, A.: *Alan Turing – The Enigma of Intelligence.* London: Unwin Paperbacks 1985.
[12] Horáková, J.: Myslet interaktivitu. In: *Entermultimediale 2*. Praha: CIANT, Mezinárodní centrum pro umění a nové technologie 2005, s. 69-70.
[13] Horáková, J., Kelemen, J.: From golem to cyborg – a note on the cultural evolution of the concept of robots. *Human Affairs* **16** (2006) 83-98.
[14] Huhtamo, E.: It is interactive – but is it art? In: *Proc. Computer Graphics Visual Processing*. New York: ACM 1993, s. 133-135.
[15] Chalmers, D.: *The Conscious Mind.* Oxford: Oxford University Press 1996.
[16] Kelemen, J.: *Postmoderný stroj.* Bratislava: F. R. & G. 1998.
[17] Kelemen, J.: *Kybergolem*. Olomouc: Votobia 2001.
[18] Kelemen, J.: O umení nechať žiť – o digitálnej metatvorbe. *Profil súčasného výtvarného umenia* **11**, č. 3-4 (2005a) 32-52.

[30] Více o této tématice lze najít v [17] a v [23].

[19] Kelemen, J.: On computational study of embodiment – some remarks and an example. *Computing and Informatics* **24** (2005b) 617-627.
[20] Krueger. M. W.: Responsive environments. In: *The New Media Reader* (N. Wardish-Fruin and N. Montfort, eds.). Cambridge, Mass.: The MIT Press 2003, pp. 379-389.
[21] Kubík, A.: *Inteligentní agenty*. Brno: Computer Press 2004.
[22] Langton, C. G.: Artificial life. In: *Artificial Life, Proc. Interdisciplinary Workshop on Synthesis and Simulation of Living Systems* (C. G. Langton, Ed.). Redwood City, Cal.: Addison-Wesley 1989, s. 1-47.
[23] Lévy, P.: *Kyberkultura.* Praha: Karolinum 2000.
[24] Lister, M., Dovey, J., Giddings, S., Grant, I., Kelly, K.: *New Media: A Critical Introduction.* New York: Routledge 2003.
[25] Lúčny, A.: Building complex systems with agent-space architecture. *Computing and Informatics* **23** (2004) 1-36.
[26] Manovich, L.: *The Language of New Media.* Cambridge, Mass.: The MIT Press 2002
[27] Markoš, A., Kelemen, J.: *Berušky, andělé a stroje*. Praha: Dokořán 2004.
[28] Minsky, M.: K-lines – a theory of memory. *Cognitive Science* **4** (1980) 117-133.
[29] Minsky, M.: *The Society of Mind*. New York: Simon and Schuster 1986.
[30] Minsky, M.: *Konštrukcia mysle.* Bratislava: Archa 1996.
[31] Minsky, M.: *The Emotion Machine*. New York: Simon & Schuster 2006.
[32] Naimark, M.: Interactive art – maybe it's a bad idea. In: *Cyberarts International Compendium Prix Ars Electronica*. Vienna and New York: Springer 1998. Dostupné rovněž na http://www.naimark.net/writing/badidea.html (1.3.2007)
[33] Penrose, R.: *Makrosvět, mikrosvět a lidská mysl.* Praha: Mladá fronta 1999.
[34] Rusnáková, K.: *V toku pohyblivých obrazov.* Bratislava: Afad Press 2005.
[35] Searle, J.: *Mysl, mozek a věda.* Praha: Mladá fronta 1994.
[36] Schröter, M. (Ed.): *Briefe am Wilhelm Fliess*. Frankfurt am Main: Fischer Verlag 1986.
[37] Thagard, P.: *Úvod do kognitivní vědy.* Praha: Portál 2001.
[38] Turing, A. M.: On computable numbers, with an application to the Entscheidungsproblem. *Proc. London Mathematical Society* **42** (1936) 230-265; corrections in **43** (1937) 544-546.
[39] Turing, A. M.: Computing machinery and intelligence. *Mind* **59** (1950) 433-460.
[40] Turner, V.: *The Ritual Process.* Chicago, Ill.: Aldina 1966.
[41] Virilio, P.: *Stroj videnia.* Bratislava: Camera lucida 2002.
[42] Warwick, K.: *I, Cyborg.* Urbana, Ill.: The University of Illinois Press 2004.
[43] Whitelaw, M.: *Metacreation – Art and Artificial Life.* Cambridge, Mass.: The MIT Press 2004.

Intencionalita zážitkov

Juraj Hvorecký[1]

Abstrakt. Delenie mentálnych stavov na propozičné postoje a zážitky čiastočne parazituje na odlišnom ponímaní intencionality u oboch druhov stavov. Kým intencionalita presvedčení a túžob je zrejmá, zameranosť niektorých zážitkov na predmety je diskutovateľná. V texte sa zaoberáme viacerými stratégiami hľadania intencionality nálad a zisťujeme problémy každej z dominujúcich teórií.

Štandardné úvahy o mysli delia mentálne entity na dve základné kategórie. Na jednej strane stoja propozičné postoje, ktoré najjednoduchšie spoznáme podľa tradičnej „že“ formy a prítomnosti propozičného obsahu. Paradigmatickými príkladmi propozičných postojov sú presvedčenia a túžby. Na strane druhej je heterogénnejšia skupina, označovaná všeobecne ako zážitky. Spadajú do nej natoľko odlišné javy, že by sa táto množina dala poľahky označiť ako doplnková, charakterizovaná nepropozičnosťou svojich prvkov. Zaraďujeme do nej senzorické vnemy (predstavy, následné obrazy), pocity (svrbenia, bolesti) a emócie (hnev, strach) [1], str. 59. Rôznorodosť zážitkov spôsobuje filozofii mysle a kognitívnym vedám ťažkosti. Ťažkosti nie sú vyvolané iba rôznorodými účinkami zážitkov na organizmus, ale aj ich nejasnou vnútornou štruktúrou.

Štruktúra propozičných postojov sa zdá byť jasná. Každý propozičný postoj sa skladá z propozície a z módu vzťahovania sa k tejto propozícií. Myšlienka, že Bush vyhrá voľby, sa skladá z propozície „Bush vyhrá voľby“, ku ktorej sa vzťahujeme módom myslenia. K rovnakej propozícii sa dokážeme vzťahovať aj inými módmi, napríklad túžbou. Výsledkom je túžba, že Bush vyhrá voľby. Táto schéma vzťahovania sa označuje ako *intencionálna*. Intencionalitou rozumieme každý taký vzťah, kde jedna entita zastupuje druhú. Inak povedané, všade tam, kde jedna entita je *o* druhej entite. Každého hneď napadne, že intencionalita je pomerne rozšírená, pretože zastupujúcu úlohu zohráva vo svete mnoho entít. Obrazy, jazykové výrazy, predstavy alebo presvedčenia niečo zastupujú, inak by neboli tým, čím sú. Napriek rozšírenosti vzťahu zastupovania je jeho uvedenie v súvislosti s mentálnymi entitami namieste. Existujú dobré dôvody sa nazdávať, že intencionalita je primárne mentálnou záležitosťou. Aj keď dejiny filozofie zaznamenali silné argumenty v prospech primárnosti jazykovej intencionality ([13], najmä kapitola 15), dnes sa za nositeľa primárnej intencionality považujú mentálne stavy. Hoci argumentácia v prospech primárnosti mentálnej intencionality nie je ťažiskom tohto článku, v stručnosti ju pripomeniem.

Jazykové výrazy nemôžu byť primárnymi nositeľmi intencionality, pretože niečo musí zaručovať správnosť ich aplikácie na konkrétne situácie ich použitia. Správnu, ale nesystematickú aplikáciu jazykových výrazov by sme za jazyk nepovažovali. Jazyk aplikuje správne iba ak ho zamýšľame použiť vhodným spôsobom. Lenže toto zamýšľanie je samo intencionálne. Jeho intencionalita však nemôže byť jazyková, pretože by vyžadovala ďalšie zamýšľanie vyššieho rádu. Ak chceme zastaviť hroziacu špirálu odkazovania na stále vyššie rády, musíme pripustiť existenciu intencionality na

[1] Filosofický ústav Akademie věd ČR, E-mail: juraj@hvorecky.com

nejazykovej, mentálnej úrovni. Ostatné formy intencionality budeme považovať za sekundárne, parazitujúce na primárnej intencionalite mentálneho (detailnejšie sa problému venuje napr. [6]). O problematickosti jazykovej intencionality svedčí aj častý fenomén neschopnosti vyjadriť vlastné myšlienky. Ak by primárnosť intencionality niesol jazyk, potom by neexistoval spôsob ako daný jav vysvetliť. Ak je však intencionalita prvotne daná mentálnej sfére, záhada neschopnosti vyjadriť myšlienky zmizne ([9], str. 277).

Otázka, ktorá bude stredobodom našej pozornosti, sleduje intencionalitu nepropozičných mentálnych stavov. Tento záujem nie je samoúčelný. Od čias Franza Brentana z konca 19. storočia sa v sociálnych vedách traduje demarkácia mentálnych stavov založená na intencionalite. Pri snahe založiť psychológiu ako vedu a odlíšiť ju od fyziológie sa Brentano oprel o intencionalitu. Kým fyziologické javy nie sú o ničom, mentálne javy sú, podľa neho, vždy o niečom. Intencionalita slúži ako pohodlné kritérium rozlíšenia predmetu psychológie od predmetov ostatných vied o človeku. Brentano prirodzene predpokladal, že zážitky sú intencionálne: „pri vneme niečo vnímame, pri súde pripúšťame alebo odmietame, v láske milujeme, v nenávisti nenávidíme, v túžbe po niečom túžime, atď.“ [3], str. 88. Brentanova téza má nepochybne obrovskú explanačnú silu. Jednoduchá vlastnosť elegantne rieši inak zložitú otázku kategorizácie mentálnych fenoménov. Otázkou ale zostáva, či elegancia zodpovedá skutočnosti. Sú naozaj všetky mentálne stavy intencionálne?

Dvadsiate storočie je plné argumentov, ktoré majú podporiť aj vyvrátiť obe možnosti. Na jednej strane stoja odporcovia Brentanovej tézy, ktorí tvrdia, že takýto prístup nie je verný fenomenológii zážitkovej skúsenosti. Na to odpovedajú jej zástancovia a snažia sa ukázať, že elegancia Brentanovho tvrdenia sa dá zachrániť a stojí za to o ňu bojovať. Odporcovia intencionalizmu sa opierajú o celkom legitímny argument. Je pravdou, že Brentanom uvádzané príklady mentálnych stavov sú intencionálne. Avšak existuje celá skupina iných mentálnych stavov, ktoré na nič neodkazujú. John Searle je jedným z prvých mysliteľov, ktorý na túto skutočnosť upozornil: „jednoducho neplatí, že by všetky naše mentálne stavy boli takto zamerané, čiže intencionálne. Ak mám bolesti, žialim, niekto ma šteklí alebo ma čosi svrbí, odpovedajúce stavy vedomia nie sú na nič zamerané“ [12], str. 74. Všimnime si, že Searle uvádza dva druhy mentálnych stavov, ktoré podľa neho odporujú Brentanovej téze. Jednak sú to telesné pocity, ktoré vstupujú do mentálnej sféry (šteklenie, svrbenie, bolesť) a zvláštnym spôsobom ju modifikujú. Do tejto skupiny však nepatrí žiaľ, ktorý by sme zrejme najsprávnejšie označili ako náladu. Iní autori považujú za najproblematickejšiu skupinu mentálnych stavov práve nálady. Baierová uvádza medzi problematickými javmi, ktorých intencionalita nie je jasná, stavy depresie, vážnosti, popudlivosti, eufórie alebo útočnosti [2], str. 3. Roger Lamb, citujúc Thalberga, hovorí podobne o depresii, eufórii a úplnej apatii [10], str. 107.

Dôvody ťažkostí pri kategorizovaní uvedených javov sú zrejmé. Kým propozičné postoje neproblematicky odkazujú k propozíciám a aj mnohé zážitky a emócie sa zdajú byť o svojich objektoch, určiť objekt, ku ktorému sa vzťahujú nálady a telesné stavy, nie je jednoduché. Pri hľadaní intencionálnych objektov telesných stavov sa môžeme oprieť aspoň o ich spojitosť s telom. Keď máme bolesti, bolí nás niečo niekde. Lokalizácia telesných pocitov je často jednoznačná. Nálady nám však nijakú pomoc v tomto smere neposkytujú. Ani s telesnosťou mentálnych stavov však nie je všetko úplne v poriadku. Fantómové bolesti u pacientov s amputovanými končatinami sú veľmi dobre zdokumentované a ich rozšírenie je takmer univerzálne (pozri [4]). S ich

intencionálnosťou je problém. Ako môžu odkazovať na časť tela, ktorá nejestvuje? S náladami je to ešte horšie. Často síce vieme určiť pôvodcu nálady, ale ten nemusí odpovedať požadovanému intencionálnemu objektu. Keď vo mne ráno suseda vyvolala mrzutosť, obraciam ju na všetkých a nedá sa povedať, že by predmetom mrzutosti bola suseda alebo jej správanie. Mrzutosť zasahuje mentálne správanie globálne, ciele si nevyberá a do jej spár poľahky spadne aj niekto, koho máme inak v obľube. Rovnaké pôsobenie vyvolávajú aj iné nálady. Dominujú ostatným pochodom mysle a málokedy sa dajú ovplyvniť. Nepodriadiť ich pôsobeniu vybrané mentálne obsahy nie je jednoduchá kognitívna úloha. Odporcovia Brentanovej tézy preto tvrdia, že z fenomenologického hľadiska je nespojitosť zážitkov a nálad s intencionálnymi objektmi zrejmá. Táto kritika sa na Brentana a jeho nasledovníkov znáša napriek tomu, že sám je za predchodcu filozofickej fenomenológie považovaný.

Priame fenomenologické závery nebránia priaznivcom tézy o intencionalite ako poznávacom znaku mentálneho v obhajobe ich elegantného demarkačného kritéria. Intencionalisti (dnes často vystupujúci pod názvom „reprezentacionalisti“) sa pokúšajú nájsť spôsoby, akými Brentanovu tézu zachrániť. Ponúkajú argumenty, ktoré demonštrujú intencionálnosť telesných pocitov a psychických nálad. Ich riešenia ako zahrnúť do všeobecnej intencionálnej schémy telesné pocity sme už naznačili. Prípady bolesti alebo svrbenia sú celkom presne lokalizovateľné a majú úzku spojitosť s telom. Podľa reprezentacionalistov je intencionálnym obsahom takýchto zážitkov jednoducho stav tela. Bolesť indikuje porušenie tkaniva, svrbenie jeho podráždenie [14], kap. 4. Predstavujú signál organizmu, že s telesnou schránkou nie je niečo v poriadku. Ako každý signál, aj tento je intencionálny. Prekážkou postupu intencionalistov nie sú ani fantómové bolesti. Tie zohrávajú rovnakú signalizačnú úlohu, akurát to robia k neexistujúcemu predmetu. To ale nepredstavuje veľkú prekážku, pretože vzťahovanie sa k neexistujúcim entitám je bežné aj u paradigmatických prípadov intencionality, akými sú presvedčenia a túžby. Niektorí ľudia sú presvedčení, že im pomáhajú homeopatické lieky, iní túžia žiť po smrti. Aj keby homeopatiká nikdy nikomu nepomohli a posmrtný život zostal snom, oba fiktívne javy zostanú naďalej objektmi ľudských myšlienok a túžob. Ako je vzťahovanie sa k nejestvujúcim javom možné je predmetom veľkých debát, ale tie stoja mimo rámec našej diskusie. Postihujú totiž rovnako propozičné postoje aj zážitky.

Intuície o fenomenálnej povahe nálad určujú stratégiu reprezentacionalistov pri obhajobe Brentanovej tézy v oblasti nálad. Zaznamenali sme, že nálady sú orientované globálne a postihujú nás svojou celostnosťou. John Haugeland sa vyjadril, že nálady sú ako „výpary, ktoré preniknú a napadnú všetko, čím sme“ [7], str. 235. Depresia alebo eufória nás pohlcujú úplne a menia náš vzťah k ostatným procesom mysle. Práve globálny charakter smerovania nálad sa zástancovia Brentana snažia využiť. Ako píše Kenny: „často nás postihujú depresie, ktorým nevieme odolať a celé dni všetko vyzerá temne; ale neopodstatnená depresia nie je nepredmetná, predmetmi depresie sú veci, ktoré sa javia temne“ [8], str. 61. Podobne postupuje aj Solomon, keď hovorí, že „nálady… sú niekedy o celom svete“ [11], str. 133. Analogické prístupy, hoci stoja na fenomenálnej skúsenosti, nie sú príliš zrozumiteľné. Špecifikácia intencionálneho objektu je v nich príliš nejasná. Byť o celom svete alebo o jeho temných stránkach nie je veľmi ilustratívny opis zložitej situácie. Jej intencionálna povaha je prinajmenšom sporná. Hoci nálady vskutku vykresľujú celok sveta istým spôsobom, možno pochybovať, či sa tak deje, pretože svet je ich predmetom. Rozumnejšie sa zdá predpokladať, že nálady majú dôsledky, ktoré

ovplyvňujú mentálne procesy globálne a dodávajú im istý náboj. Lenže dôsledky mentálnych stavov a ich objekty sú dve rozličné veci. Lamb dokonca vyjadruje názor, že spomínané snahy vtesnať nálady do domény intencionálneho sú len „pokusmi zachovať teoretickú stavbu“ intencionalizmu [10], str. 114.

Proti sebe stoja dve silné tvrdenia. Na jednej strane je zásada dôsledne sa pridržiavať skúsenostných dát a tá vedie k odmietaniu Brentana. Neponúka však nijaký alternatívny názor na legitímnosť mentalistickej kategorizácie. Mentálne javy nemajú nijakého spoločného menovateľa a založenie psychológie ako vedy je filozoficky podozrivé. Na strane druhej stojí vernosť téze o intencionalite ako rozlišovaciemu znaku psychologických objektov. Niekedy stojí za to sa držať silnej teórie zubami-nechtami, nemalo by sa tak ale diať v príkrom rozpore s empíriou.

Keďže jednoduchá dichotómia intencionálne – neintencionálne neprináša očakávaný efekt, otvára sa pole pre alternatívne prístupy, ktoré sa intencionalitu zážitkov snažia vysvetliť komplexnejšie. Baierová [2], ovplyvnená darwinistickým čítaním Freuda, rozlišuje u emócii niekoľko vrstiev intencionálneho objektu. Okrem bezprostredného objektu, ktorým v prípade lásky je partner(ka), existuje hlboký objekt (pre freudistov matka), prípadne formálny objekt (obľúbená osoba), s ktorými musí bezprostredný objekt nejako korešpondovať. Ak si aj odmyslíme ťažkosti so špecifikovaním relácie korešpondencie, Baierovej pokus považujeme za neudržateľný. Nezdá sa totiž, že by jej teória bola obhájiteľná nezávisle od psychoanalytickej teórie, ktorá sama je zdiskreditovaná.

Zaujímavý evolučný pokus vysvetliť fungovanie emócii nachádzame u S. Cunninghamovej. Za paradigmatický biologický príklad jej slúži strach. Podľa nej sa v strachu, povedzme z hadov, odzrkadľuje reflexívna úroveň nášho vzťahovania sa k obávaným objektom, ktorá je v štandardných teóriách intencionality opomínaná. „Strach z hadov nie je len o možných hadoch v lese. Práve tak je aj o možnom vzťahu hada k nám a k nášmu pocitu blaha“ [5], str. 458. Opodstatnenosť takejto analýzy je veľmi otázna. Čo sa javí ako zaujímavé rozšírenie intencionality tejto emócie sa vo svetle bližšej analýzy ukáže ako nešťastná duplikácia problému. Stretnutie s hadom alebo obava z takéhoto stretnutia sa bezpochyby viaže na subjektívny pocit spokojnosti a jeho možné ohrozenie. Hodnotenie ohrozenia však dozaista nie je súčasťou pocitu strachu a zrejme sa spája s kognitívnymi procesmi, ktoré strachu predchádzajú, prípadne ho doprevádzajú. K nim možno zaradiť napríklad presvedčenia o hrozbe zo strany hadov alebo skúsenosti z predchádzajúcich stretnutí s hadmi. Vzťah hada k nám a nášmu blahu nie je predmetom strachu, ale strachom samotným! Zdá sa nám, že Cunninghamová zmiešava kategórie módu a predmetu. Intencionálna štruktúra strachu z hadov je zrejmá: predmetom strachu je had. Akékoľvek relačné štruktúry, ktoré sa pri styku ľudí s hadmi vyskytujú, nepatria medzi predmety intencionálnych vzťahov, ale medzi samotné mentálne stavy.

Je žiaduce znovu oživiť debaty o povahe intencionality nepropozičných mentálnych stavov, pretože doterajšie diskusie buď sťažujú demarkáciu psychológie, alebo ponúkajú intencionálne analýzy, ktoré nespĺňajú skúsenostné a formálne zásady.

Literatúra

[1] Addis, L.: Pains and Other Secondary Mental Entities. *Philosophy and Phenomenological Research* **47** (1986) 59-74.
[2] Baier, A: What Emotions Are About. *Philosophical Perspectives* 4, Action Theory and Philosophy of Mind, 1990, 1-29.
[3] Brentano, F., *Psychology from an Empirical Standpoint*, preklad A.C. Rancurello, D.B. Terrell, a L. McAlister, London: Routledge 1973.
[4] Cole J., Paillard, J.: Living without Touch and Peripheral Information about Body Position and Movement: Studies with Deafferented Subjects. In J.L. Bermúdez, A. Marcel a N. Eilan (eds.), *The Body and The Self.* Cambridge, MA: MIT Press 1995.
[5] Cunningham, S.: Two Faces of Intentionalit*., Philosophy of Science* **64** (1997) 445-460.
[6] Fodor, J.: *A Theory of Content*, Cambridge, MA: MIT Press 1990.
[7] Haugeland, J.: *Artificial Inteligence: The Very Idea*. Cambridge, MA: MIT Press 1985.
[8] Kenny, A.: *Action, Emotion and Will*. London: Routledge & Kegan Paul 1963.
[9] Knowles, M.: Some Remarks on the Intentionality of Thought. *Philosophy and Phenomenological Research* **42** (1981) 267-279.
[10] Lamb, R.: Objectless Emotions. *Philosophy and Phenomenological Research* **48** (1987) 107-117.
[11] Solomon, R.C.: *The Passions*. New York: Anchor Press 1976.
[12] Searle, J.: What Is an Intentional State? *Mind* 88 (1979) 74-92.
[13] Sellars, W.: *Empiricism and the Philosophy of Mind.* Cambridge, MA: Harvard University Press 1997.
[14] Tye, M.: *Ten Problems of Consciousness*. Cambridge, MA: MIT Press 1995.

O vzťahu prírodných a spoločenských vied

Ladislav Kvasz[1]

Abstrakt. Stať sa usiluje zasadiť konflikt medzi prírodnými a spoločenskými vedami do širšieho historického kontextu. Prináša teoretickú analýzu posunov hraníc medzi vednými disciplínami, ku ktorým dochádza počas vedeckej revolúcie. Okrem *paradigmatických disciplín* uvažuje ešte *zmiešané disciplíny*, *metaforickú oblasť* paradigmy a *eluzívnu oblasť* paradigmy. Na základe takto zjemnenej analýzy ponúka epistemologickú alternatívu ku Kuhnovej čisto sociologickej teórii vedeckých revolúcií. Namiesto ako konflikt medzi zástancami a oponentmi novej paradigmy, vedeckú revolúciu interpretuje ako konflikt medzi metódami *zmiešaných disciplín* a *metaforickou oblasťou* starej paradigmy. Z tohto hľadiska patria kognitívne vedy medzi *zmiešané disciplíny* súčasnej biologickej paradigmy a ako také môžu zohrať kľúčovú rolu pri konštitúcii novej paradigmy spoločenských vied.

1 Úvod

Medzi metódami prírodných a spoločenských vied existuje už dlhú dobu určité napätie. Ako píše T. Kuhn v knihe *Štruktúra vedeckých revolúcií* ([3], str. 30), bolo to práve toto napätie, ktoré ho viedlo k vytvoreniu pojmu paradigmy. Podľa Kuhna rozdiel medzi prírodnými a spoločenskými vedami spočíva v tom, že kým v prírodovedných disciplínach máme dočinenia s normálnou vedou založenou na všeobecne akceptovanej paradigme, v spoločenských vedách paradigma chýba a vedci neustále spochybňujú samotné základy svojich disciplín. Kuhn takto poukázal na zásadný problém. Keď však chceme tomuto problému lepšie porozumieť, ukazuje sa ako účelné pozrieť sa na napätie medzi prírodnými a spoločenskými vedami v historickej perspektíve. Cieľom predkladanej state je zasadiť konflikt prírodných a spoločenských vied do širšieho historického rámca a ukázať, že tento konflikt dnes už do veľkej miery patrí minulosti. Zlom, ktorý v minulosti jasne oddeľoval vedy o prírode od vied o kultúre, sa rozpadol na celú kaskádu stupňovitých prechodov.

Prvá vec, na ktorú pri historickom pohľade narazíme, je zistenie, že konflikt prírodných a spoločenských vied tu nebol odjakživa. V antike neexistoval rozpor medzi poznávaním človeka a spoločnosti na jednej strane a poznávaním prírody na strane druhej. To, samozrejme, neznamená, že by v antike celé poznanie tvorilo jediný harmonický celok. Aj v antike existovalo napätie analogické napätiu medzi prírodovedným a spoločenskovedným poznaním. Hranica, pozdĺž ktorej sa toto napätie manifestovalo, však prebiehala inde. Neoddeľovala prírodovedné poznanie od spoločenskovedného, ale matematické poznanie (založené na deduktívnej metóde) od „organického“ poznania (založeného na kategóriách ako účel, cieľ, pôsobenie). V tejto druhej skupine sa spolu stretli ako biologické, tak aj spoločenskovedné disciplíny, teda disciplíny, ktoré podľa nášho chápania ležia na opačných stranách barikády oddeľujúcej prírodovedné poznanie od poznania spoločenskovedného. Gréci pristupovali rovnako k poznávaniu „*vzniku*

[1] Katedra algebry, geometrie a didaktiky matematiky FMFI UK, Mlynská dolina, 842 48 Bratislava, E-mail: kvasz@fmph.uniba.sk

a zániku živočíchov“ ako k poznávaniu „*duše*“. Až v novoveku bola náuka „*o vzniku a zániku živočíchov*“ presunutá do oblasti prírodných vied, kým náuka „*o duši*“ sa stala jedným z kryštalizačných jadier novovznikajúcich spoločenských vied. Prvou úlohou tejto state je navrhnúť teoretický rámec, ktorý by umožnil racionálne rekonštruovať podobné presuny v klasifikácii vedeckých disciplín. Pokúsime sa ukázať, že to, na ktorej strane barikády oddeľujúcej „tvrdé“ disciplíny od „mäkkých“ disciplín sa určitá teória ocitne, nie je náhodné; práve naopak, týmto presunom možno porozumieť. Vyžaduje si to však určité zjemnenie Kuhnovho dichotomického delenia disciplín na základe ich vzťahu k paradigme.

2 Klasifikácia vedeckých disciplín na základe ich vzťahu k paradigme

Aby sme mohli rekonštruovať presuny v klasifikácii vedeckých disciplín na „tvrdé“ a „mäkké“, musíme si vytvoriť jasnejšiu predstavu o „topografii vedeckej krajiny“ ležiacej medzi týmito pólmi. Ako prvý krok navrhujeme opustiť *terminológiu* delenia disciplín na „tvrdé“ a „mäkké“. „Tvrdé“ disciplíny budeme označovať termínom ***paradigmatické disciplíny***. Na presnejšiu charakterizáciu určitej časti „mäkkých“ disciplín navrhujeme zaviesť termín ***eluzívna oblasť paradigmy*** (oblasť unikajúca paradigmatickému uchopeniu). Okrem týchto dvoch druhov disciplín, ktorých rozbor ovláda úvahy o metodológii spoločenských vied, navrhujeme zaviesť ešte dva prechodné typy, ležiace medzi paradigmatickou oblasťou a eluzívnou oblasťou paradigmy.[2] Ako sa ukáže, najzaujímavejšie zmeny sa odohrávajú práve v tomto priestore.

Prvú kategóriu vedeckých disciplín, ležiacich v priestore medzi paradigmatickou oblasťou a eluzívnou oblasťou paradigmy, navrhujeme označiť termínom ***zmiešané disciplíny***. Tento termín sa používa v histórii vedy na označenie pozoruhodnej spleti disciplín zo sklonku antiky, ako sú *Ptolemaiova astronómia, Euklidova optika, Archimedova teória páky, Archimedova teória stability plávajúcich telies, teória jednoduchých strojov* (v antike označovaná termínom *mechanika* a pripisovaná Aristotelovi). Uvedené disciplíny majú spoločné to, že sa v nich exaktný matematický jazyk používa na výklad javov, ktoré podľa antického chápania nie sú matematizovateľné, lebo na nich má podiel látka. Podľa aristotelovskej metodológie by sa matematika pri opise týchto javov nemala používať. Uvedené teórie nie sú deduktívne, nespĺňajú štandardy matematického poznania, ale na druhej strane nie sú vybudované ani kauzálne, ako má podľa Aristotelových predstáv vyzerať výklad javov v eluzívnej oblasti paradigmy. Teda v celom rade prípadov sa prax antickej vedy nedrží štandardov, ktoré metodológii uložil Aristoteles, a vytvára disciplíny, ktorých metodologický status je pochybný. To, že páka, zrkadlo, teleso ponorené do kvapaliny či kladkostroj sú materiálne objekty, a napriek tomu sa pri ich opise používa matematický aparát, je z hľadiska antického chápania vedeckého poznania škandál. Zmiešané disciplíny pritiahli pozornosť historikov vedy, lebo počas vedeckej revolúcie 17. storočia nadobudli veľký význam. Galileo napísal spis o plávaní telies, Fermat a Descartes sa zaoberali otázkou lomu svetla, Galileo dosiahol závažné objavy v astronómii a teórii jednoduchých strojov. Možno povedať, že to boli práve zmiešané

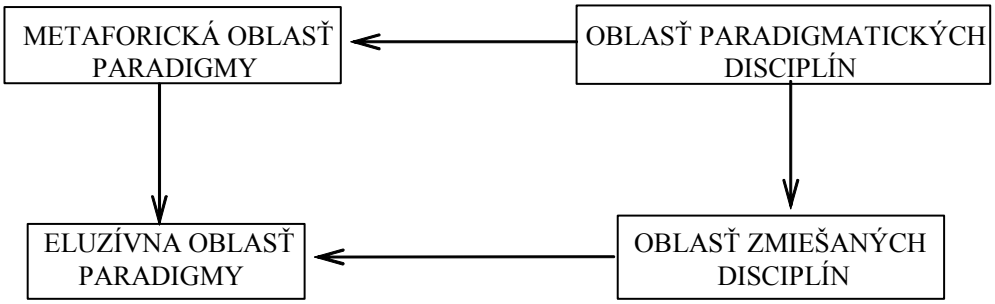
[2] Kuhnov pojem paradigmy je viacznačný, pričom možno rozlíšiť tri druhy paradigiem – paradigmu *idealizácie*, paradigmu *reprezentácie* a paradigmu *objektácie* ([5], str. 230). Z hľadiska predkladanej state je dôležitá idealizácia, preto všade v ďalšom texte pod paradigmou budeme rozumieť paradigmu idealizácie.

disciplíny, v ktorých sa rodili základné pojmy novej paradigmy.

Druhú dodatočnú oblasť disciplín, ležiacich v priestore medzi paradigmatickou oblasťou a eluzívnou oblasťou paradigmy, navrhujeme označiť termínom ***metaforická oblasť paradigmy***. Je to oblasť, ktorá predstavuje protipól oblasti zmiešaných disciplín. Kým v prípade zmiešaných disciplín sa pojmy a metódy paradigmy používajú v presnom a jednoznačnom zmysle a problém je len v tom, že sa používajú za hranicami oblasti, v ktorej je ich použitie metodologicky obhájiteľné, v metaforickej oblasti sa základné pojmy paradigmy používajú v prenesenom, často veľmi nepresnom a iba približnom zmysle. Príkladom metaforickej oblasti antickej paradigmy je *Aristotelova teória miestneho pohybu*, podľa ktorej ťažké telesá padajú dole a ľahké stúpajú nahor. V stati venovanej výkladu karteziánskej fyziky ([10], str. 228-229) sme sa snažili ukázať, že Aristotelova teória miestneho pohybu je geometrická teória. Základom Aristotelovho opisu je totiž predstava geometricky usporiadaného univerza a pohyb sa chápe ako prechod medzi rôznymi miestami tohto geometrického usporiadania. Dôležité je si však uvedomiť, že geometria (t.j. paradigmatická disciplína antickej vedy) sa tu používa inak než v prípade zmiešaných disciplín. Geometria v Aristotelovej predstave kozmu nefiguruje explicitne ako súbor exaktných pojmov a metód na robenie konštrukcií a dokazovanie poučiek (ako geometriu používa vo svojej teórii rovnováhy na páke Archimedes), ale implicitne, ako súbor metafor, pomocou ktorých nachádzame v neprehľadnej spleti javov poriadok a zmysel. Preto napriek tomu, že Aristotelovo chápanie pohybu je organické a matematickému opisu sa vymyká (patrí do eluzívnej oblasti paradigmy), jeho časť – teória miestneho pohybu – je opísaná pomocou geometrických metafor.

Vidíme, že okrem *paradigmatickej oblasti*, teda oblasti uchopenej prostriedkami paradigmy v zhode s jej metodologickými normami, a *eluzívnej oblasti*, teda oblasti, ktorá sa paradigmatickému uchopeniu vymyká, umožňuje paradigma sformovať ďalšie dve oblasti vedeckých disciplín. Jednak poskytuje technický aparát na sformovanie *oblasti zmiešaných disciplín*, ktoré síce používajú exaktné pojmy a postupy paradigmy, ale aplikujú ich na javy, pre ktoré neboli zamýšľané, a kde preto nie je možné splniť metodologické štandardy diktované paradigmou. Okrem toho paradigma umožňuje sformovať *metaforickú oblasť*, ktorú tvoria javy natoľko zložité, že na presné technické použitie pojmov a metód paradigmy nemožno ani pomyslieť. Napriek tomu však paradigma poskytuje celý rad metafor, ktoré umožnia aspoň kvalitatívne týmto javom porozumieť, a tak ich pričleniť na okraj racionálneho diskurzu vytvoreného paradigmou. Ak chceme porozumieť vzťahu prírodovedných disciplín (tvoriacich paradigmatickú oblasť novovekej vedy) a spoločenskovedných disciplín (ležiacich v eluzívnej oblasti tejto paradigmy), je vhodnejšie obraz vedy upresniť a namiesto opozície „tvrdých“ a „mäkkých“ disciplín položiť nasledovnú schému:[3]

[3] Je to takmer pravidlo, že všade tam, kde sme konfrontovaní s určitým nezmieriteľným protikladom, je tento protiklad dôsledkom zjednodušeného obrazu skutočnosti, obrazu, ktorý ignoruje medzipolohy.



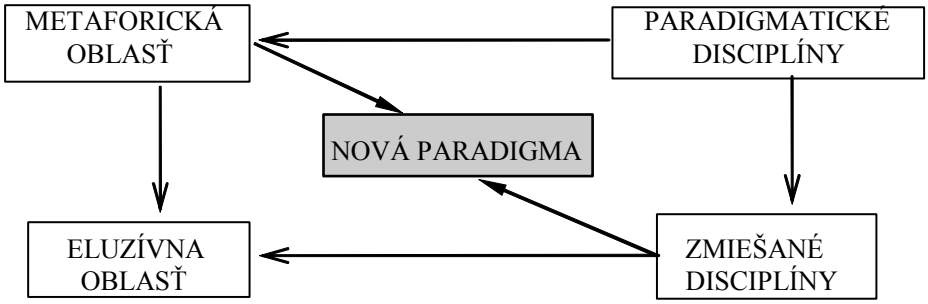
3 Náčrt rekonštrukcie vedeckej revolúcie 17. storočia

Uvedená schéma umožňuje oslabiť protiklad medzi paradigmatickou oblasťou, tvorenou „tvrdými“ disciplínami, a eluzívnou oblasťou paradigmy, ktorú tvoria „mäkké“ disciplíny, a tak otvára možnosť racionálnej rekonštrukcie vedeckej revolúcie 17. storočia. Ukazuje sa, že práve zmiešané disciplíny a ich konflikt s metaforickou oblasťou paradigmy boli jej motorom. Novoveká fyzikálna paradigma sa nezrodila vnútri paradigmatickej oblasti. Paradigmatickou disciplínou antickej vedy bola matematika a zrod fyziky síce podnietil vznik viacerých nových matematických disciplín, ale o tom, že by v matematike došlo k nejakej falzifikácii či odmietnutiu predošlého výskumu, sotva môže byť reč. Zásadne sa nezmenila ani eluzívna oblasť starej paradigmy (t.j. oblasť organického). V priebehu vedeckej revolúcie 17. storočia bola biológia na okraji vedeckého záujmu a do centra pozornosti sa dostala až na sklonku 18. storočia, kedy revolúcia bola už zavŕšená. Vedecká revolúcia 17. storočia sa udiala pri strete zmiešaných disciplín (astronómie, optiky a teórie jednoduchých strojov) a metaforickej oblasti paradigmy (geocentrického obrazu kozmu). A je to aj prirodzené.

V paradigmatickej oblasti antickej vedy, teda v matematike, sú metodologické štandardy natoľko prísne, že zásadná falzifikáciu celkového obrazu neprichádza do úvahy. Eluzívna oblasť paradigmy (teda oblasť budúcej biológie) je zas ešte málo preskúmaná, a preto v nej nemožno očakávať žiadny zásadnejší pokrok. Sú to práve zmiešané disciplíny, kde paradigma poskytuje dostatočne efektívny aparát, takže ich pokrok je pomerne rýchly, ale na druhej strane celý podnik zmiešaných disciplín je dosť pochybný, takže je tu prirodzené očakávať objavenie sa čohosi zásadne nového a nečakaného, čohosi, čo bude v rozpore s tým, na čo sme si v paradigmatickej oblasti zvykli. Metaforická oblasť paradigmy je pre vedeckú revolúciu dôležitá zase preto, lebo sa pohybuje na samom okraji toho, čo ešte paradigma umožňuje tematizovať, a preto sú v nej zafixované základné kultúrne projekcie[4] a s nimi spojený emocionálny náboj. Samotné zmiešané disciplíny by

[4] Ako píše o Aristotelovej teórii pohybu Ph. Frank: „*Teória pohybu bola veľmi dôležitá pre život človeka, lebo podporovala jeho vieru v hierarchickú štruktúru spoločnosti. Dodávala odvahu mravnému správaniu ľudských bytostí.*“ ([1], str. 18). To znamená, že eluzívna a metaforická oblasť paradigmy tvorí sociálne konštruované poznanie. Klasifikácia vedeckých disciplín na základe ich vzťahu k paradigme umožňuje zmierniť konflikt, ktorý vo filozofii vedy vyvolalo vystúpenie sociálneho konštruktivizmu. Kritici tohto prúdu oprávnene namietali, že „tvrdé fakty“ nemôžu byť sociálne konštruované. Fakty paradigmatických disciplín sú jednoznačne určené ich metódami, a tak pre sociálne konštruovanie niet priestoru. Sociálne konštruovaná je však eluzívna oblasť paradigmy. V procese, pri ktorom sa z eluzívnej oblasti odčlení

asi vedeckú revolúciu nespustili. Keby sa Galileo podriadil cirkvi a kopernikovský systém by pojal ako hypotézu, teda keby sa bol obmedzil na technickú oblasť zmiešaných disciplín a nekonfrontoval by ju s geocentrickým obrazom sveta, asi by sa podarilo astronomické objavy udržať na periférii záujmu verejnosti ako zložité a nezrozumiteľné, ale neškodné hypotézy. Dynamika vedeckej revolúcie 17. storočia je daná práve konfliktom zmiešaných disciplín a metaforickej oblasti, keď sa nie celkom zaručené výsledky vedeckého bádania[5] dostanú do konfliktu s metaforami, pomocou ktorých artikulujeme svoje postavenie vo svete.



V tejto schéme paradigmatickými disciplínami rozumieme paradigmatické disciplíny *starej* paradigmy a to isté platí aj o zmiešaných disciplínach, o metaforickej aj o eluzívnej oblasti. Keď sa zameriame na vedeckú revolúciu 17. storočia, schéma opisuje

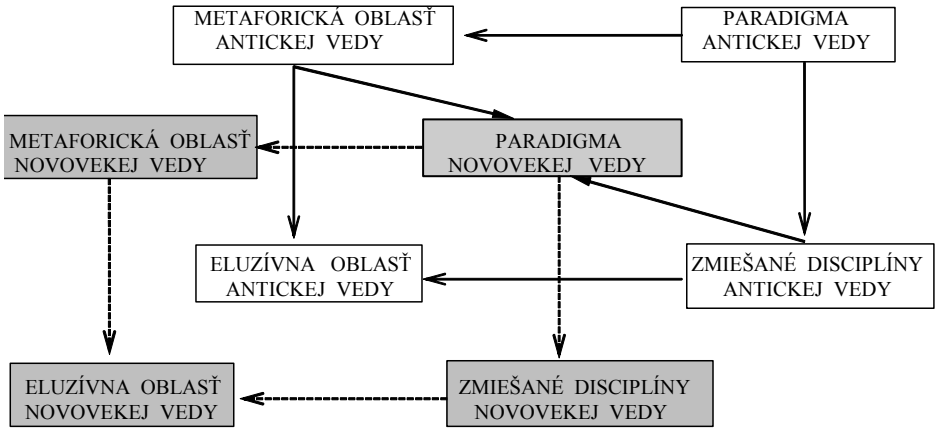
metaforická oblasť a vstúpi do tvorby novej paradigmy, možno vidieť transformáciu sociálne konštruovaného poznania na poznanie konštituované paradigmaticky. Aristotelova hierarchická predstava kozmu je sociálne konštruované poznanie, lebo prenáša hierarchický poriadok spoločnosti na opis stavby vesmíru. V priebehu vedeckej revolúcie 17. storočia sa táto predstava dostala do rozporu s astronomickými pozorovaniami a veda nahradila geocentrický obraz aristotelizmu acentrickým univerzom Newtonovej fyziky. Nová paradigma však absorbovala iba metaforickú oblasť antickej vedy, kým v jej ***eluzívnej oblasti***, predstavovanej biológiou, idea hierarchicky usporiadaného celku pretrvala ešte dve storočia, až kým darwinovská revolúcia nahradila hierarchický poriadok živých organizmov s človekom ako jeho korunou evolučne konštituovaným súborom genetických príbuzností. Eluzívna oblasť antickej vedy bola teda paradigmaticky uchopená až po dvoch krokoch *ideácie*. V prvom kroku, predstavovanom fyzikálnou revolúciou, sa rozbíja jednota aristotelovského obrazu sveta a jeho metaforická časť (teória pohybu a stavba vesmíru) je integrovaná do novej paradigmy. K plnému nahradeniu sociálne konštruovanej predstavy hierarchického poriadku paradigmaticky konštruovanou predstavou dochádza však až v druhom kroku. Preto o dva kroky idealizácie dozadu bolo paradigmatické poznanie sociálne konštruované. Sociálny konštruktivizmus má tak pravdu pokiaľ ide o pôvod poznania, ale mýli sa, pokiaľ ide o jeho obsah. Len čo sa určitá oblasť dostane do sféry paradigmatických disciplín, jej poznatky sú konštituované metódami paradigmy.

[5] Galileov ďalekohľad mal veľkú chromatickú chybu. Túto sa podarilo odstrániť až Newtonovi, keď namiesto šošovkového ďalekohľadu skonštruoval zrkadlový ďalekohľad. Okrem toho v Galileovej dobe ešte nebola vypracovaná teória umožňujúca porozumieť tomu, ako v ďalekohľade vzniká obraz. Takúto teóriu vypracoval až Kepler. Navyše nebol ešte teoreticky zvládnutý ani zákon lomu svetla, teda základný princíp, na základe ktorého ďalekohľad funguje. Vysvetlenie zákona lomu podali až Descartes a Fermat. Samozrejme, to neznižuje význam Galileových objavov, ba naopak ešte zvýrazňuje jeho geniálnu intuíciu a intelektuálnu odvahu. Ale na druhej strane nás núti uznať, že až práce Keplera, Descarta, Fermata a Newtona poskytli Galileovým pozorovaniam pevný metodologický základ, ktorý z nich spravil východisko modernej astronómie. Dnes máme sklon vidieť tieto pozorovania už od samého začiatku ako pevne metodologicky osnované, a tak nedoceňujeme problémy spojené s Galileovými objavmi.

skutočnosť, že ***paradigma novovekej fyziky sa nezrodila ani v paradigmatickej oblasti, ani v eluzívnej oblasti antickej paradigmy, ale medzi nimi***. Paradigmatickou oblasťou antickej vedy bola matematika, ktorá sa v 17. storočí síce intenzívne rozvíjala (zrodila sa analytická geometria, diferenciálny a integrálny počet), ale tento vývoj sa odohrával v medziach normálnej vedy. V matematike nikto vážne nespochybňoval jej minulé výsledky. Eluzívnou oblasťou antickej vedy bola oblasť organického a tvorcovia novovekej fyziky sa biológii skoro úplne vyhli. Galileo iba okrajovo uvažoval o otázke rozmerov organizmov a Descartes robil príležitostné pitvy. Je dôležité, aby sme si jasne uvedomili túto skutočnosť, lebo ukazuje, kde máme v súčasnosti hľadať zdroj revolučných zmien v spoločenských vedách. Pre novovekú vedu je paradigmou fyzika a eluzívnou oblasťou je oblasť subjektívneho (Descartovo *res cogitans*), teda oblasť spoločenských vied. Vyššie uvedená schéma ukazuje, že všetci, čo čakali či čakajú na „*Newtona spoločenských vied*“ (alebo si myslia že už prišiel), čakajú pri nesprávnych dverách. Spoločenské vedy tvoria eluzívnu oblasť fyzikálnej paradigmy a eluzívna oblasť nie je prístupná vedeckým metódam, a preto nebude hrať významnú úlohu v súčasných revolučných zmenách. ***Najbližší zásadný zlom sa nebude odohrávať v eluzívnej oblasti fyzikálnej paradigmy, teda v oblasti vied o ľudskom subjekte, ale na rozhraní zmiešaných disciplín a metaforickej oblasti fyziky***. Preto sa na toto rozhranie pozrieme bližšie.

4 Revolúcia v biológii

Keď si chceme utvoriť predstavu o vedeckej revolúcii, ktorá prebieha v súčasnosti, ukazuje sa ako užitočné prejsť k ešte komplexnejšej schéme, ktorá by obsahovala znázornenie zmiešanej, metaforickej a eluzívnej oblasti ako antickej paradigmy (ktorú sme opísali vyššie), tak aj novovekej fyzikálnej paradigmy, ktorej rozbor na nás teraz čaká. Ako sme už spomenuli, eluzívna oblasť antickej paradigmy, tvorená oblasťou organického, nehrala pri formovaní paradigmy novovekej vedy žiadnu úlohu. Predsa však došlo vo vzťahu s ňou k zásadnej zmene. ***Eluzívna oblasť antickej vedy sa dostala do centra priestoru medzi metaforickou oblasťou a zmiešanými disciplínami novovekej paradigmy***.



Keďže vedecká revolúcia sa odohráva práve na tomto rozhraní, znamená to, že predmetom vedeckej revolúcie, prebiehajúcej v súčasnosti, je oblasť organického, teda biológia. Biologická paradigma sa rodí z konfliktu medzi oblasťou zmiešaných disciplín a metaforickou oblasťou fyzikálnej paradigmy, preto si teraz tieto oblasti opíšeme bližšie.

4.1 Zmiešané disciplíny novovekej vedy

Zmiešané disciplíny novovekej vedy používajú aparát fyziky (ako technický aparát, teda experimentálne metódy a prístroje, tak aj teoretický aparát, teda pojmy a formalizmy) na skúmanie nefyzikálnych systémov. Pre zmiešané disciplíny je pritom charakteristické, že tento aparát používajú presne a prísne metodicky, problém je len v tom, že ho používajú na skúmanie biologických systémov, teda za hranicami fyziky. V oblasti experimentálnych techník sme svedkami spektakulárneho pokroku, vedúceho od objavu mikroskopu cez roentgenový prístroj až po počítačovú tomografiu. Ako jednu z posledných noviniek možno uviesť magnetorezonančnú tomografiu umožňujúcu vizualizáciu mozgovej aktivity pri kognitívnych procesoch [2]. Objavovanie stále nových princípov registrácie dát je zdrojom prevratných poznatkov v biológii a v medicíne. Podobne spektakulárny vývin možno sledovať aj v oblasti chemickej analýzy živej hmoty, vedúci od syntetickej výroby močoviny cez pochopenie stavby hemoglobínu až po analýzu ľudského genómu. Preto disciplíny ako biochémia, molekulárna biológia či neurofyziológia navrhujeme považovať za *zmiešané disciplíny, ktorých epistemologický status je analogický euklidovskej optike či archimedovskej teórii páky v antike*. Táto oblasť nepredstavuje problémy, snáď okrem toho, že jej filozofia vedy nevenuje dostatočnú pozornosť. Vo filozofii vedy sa vedecká teória definuje ako súbor propozícií, kým v prípade zmiešaných disciplín je ich vnútorná jednota daná skôr technickým aparátom. Zmiešané disciplíny by si rozhodne zaslúžili filozofickú reflexiu, lebo práve ony sú motorom paradigmatických zmien.

4.2 Metaforická oblasť novovekej vedy

Na rozdiel od zmiešaných disciplín je výklad metaforickej oblasti fyzikálnej paradigmy kameňom úrazu. Eluzívnou oblasťou fyzikálnej paradigmy je sféra subjektívneho. Nezáleží na tom, či ju vymedzíme ako Descartes, ktorý stál pri zrode fyzikálnej paradigmy, alebo ako Dilthey, ktorý zažil jej vyvrcholenie na sklonku 19. storočia. Dôležitý je eluzívny charakter subjektívneho. Z tejto eluzívnej oblasti sa postupne vyčlenila určitá jej časť, podobne ako sa z aristotelovskej teórie pohybu vyčlenila teóriu miestneho pohybu; t.j. časť, ktorá používa metafory z oblasti paradigmatických disciplín. Ako príklady možno uviesť *asociačnú psychológiu*, nadväzujúcu na Davida Huma, *ekonomickú teóriu obehu kapitálu*, nadväzujúcu na Francoisa Quesnayho či *sociologickú teóriu* nadväzujúcu na Augusta Comta.[6] Vo všetkých týchto disciplínách sa používajú pojmy ako proces, dynamika, rýchlosť, intenzita, sila a podobne. Avšak podľa nášho názoru ***je hovorenie o mentálnych procesoch, o intenzite psychických zážitkov či o rýchlosti asociácií analogické Aristotelovmu používaniu termínov „hore“ či „dole“ pri***

[6] Rovnako sem patrí aj historizmus, teda presvedčenie, že v ľudských dejinách (teda v eluzívnej oblasti fyzikálnej paradigmy) platia zákony, analogické fyzikálnym zákonom. Keďže historizmus neviedol ku vzniku žiadnej vednej disciplíny, jeho rozborom sa nebudeme zaoberať.

opise miestnych pohybov. Ide o to, že tieto pojmy sa nepoužívajú v ich striktnom, teda fyzikálnou paradigmou fixovanom význame. Fyzikálne procesy prebiehajú v priestore a práve metrická štruktúra priestoru umožňuje hovoriť o ich rýchlosti. Procesy asociácie predstáv však neprebiehajú v žiadnom jasne vymedzenom priestore, a tak pojem procesu je tu použitý iba v metaforickom zmysle. To nie je chyba, nám tu ide len o konštatovanie faktu. Rovnako Comteovo používanie termínov ako sociálna *statika* a sociálna *dynamika* je metaforické. V striktnom, teda fyzikálnou paradigmou fixovanom význame termín statika označuje náuku o rovnováhe síl. Sila je presne vymedzená fyzikálna veličina, ktorá sa meria v jednoznačne definovaných jednotkách ($kg.m.s^{-2}$). Naproti tomu „sily" pôsobiace v spoločnosti nemajú žiadne jednotky, v ktorých by sme ich merali, a reči o „rovnováhe síl" sú čisto metaforické. Podobne je metaforou aj pojem pracovnej sily v ekonómii. Pracovná sila nie je skutočnou silou, nedá sa merať pomocou fyzikálnych jednotiek, ktorými meriame fyzikálne sily. Rovnako pojem práce v ekonómii je iba metaforické použitie fyzikálneho pojmu práce, ktorý je exaktne definovaný ako dráhový integrál sily. V ekonómii nie je jasné, aké sily máme integrovať a po akej dráhe.

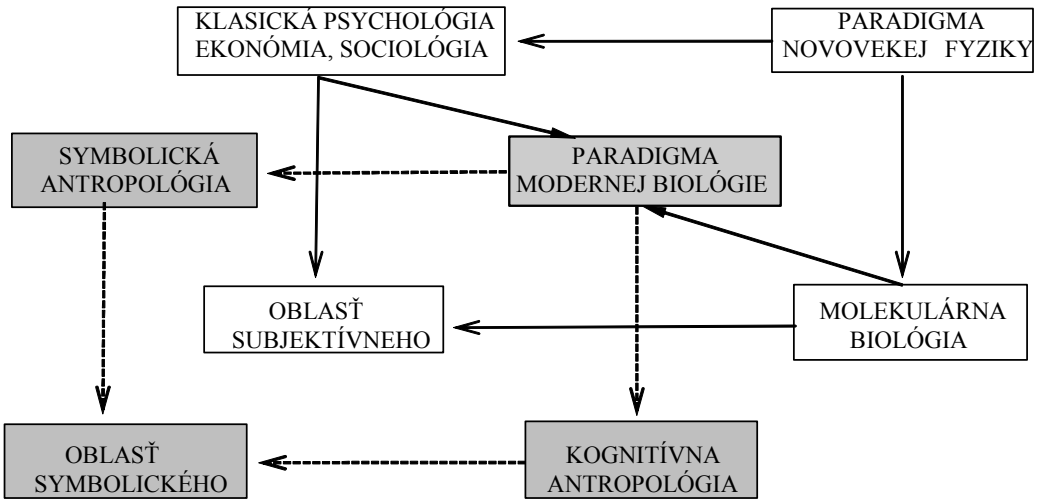
Metafory, na ktorých sa zakladajú uvedené disciplíny, nemožno konceptuálne spresniť ani jednoznačne uchopiť. Nech robíme čo robíme, zo sociálnych síl sa nedá vytvoriť pojem, ktorý by aspoň približne pripomínal svojou jednoznačnosťou pojem sily v mechanike či teórii poľa. Práve táto zásadná nepresnosť a nejednoznačnosť je príčinou toho, že tieto pojmy zahŕňame do metaforickej oblasti paradigmy. Chceme tým povedať, že disciplíny ako asociačná psychológia, politická ekonómia či pozitivistická sociológia sa snažia porozumieť svojmu predmetu pomocou metafor, ktoré preberajú z fyziky. Súčasne však predmet, na ktorý tieto metafory aplikujú, vyčleňujú z eluzívnej oblasti fyzikálnej paradigmy, teda z oblasti, ktorá sa opisu prostriedkami tejto paradigmy vymyká. To má za následok problematický status týchto disciplín, ktorý ich priam predurčuje na to, aby sa počas nasledujúcej vedeckej revolúcie zásadne zmenilo ich začlenenie.

5 Biologická revolúcia a spoločenské vedy

Slabinou úvah o rozdieloch medzi prírodnými a spoločenskými vedami je ignorovanie biológie. Všetky povzdychy nad údajnou nevedeckosťou spoločenských vied možno chápať ako artikuláciu skutočnosti, že *spoločenské vedy spadajú do eluzívnej oblasti fyzikálnej paradigmy*. Toto konštatovanie je správne, ale príklad biológie ukazuje, že fyzikálna paradigma nevyčerpáva celú oblasť prírodných vied. Medzi fyzikálnymi vedami, ktoré sa väčšinou berú ako príklad prírodných vied, a biologickými vedami existujú zásadné rozdiely v povahe empirického základu, ako aj logickej stavby ich teórií. Keď schému uvedenú v predošlej kapitole extrapolujeme o jeden krok, možno očakávať, že prelom v oblasti spoločenských vied nastane až vtedy, keď sa biologická paradigma plne konštituuje, t.j. keď vzniknú jej zmiešané disciplíny a jej metaforická oblasť. Zásadná zmena, ktorá nastane vznikom biológie ako paradigmatickej vedy, spočíva v tom, že spoločenskovedné disciplíny, ktoré tvorili eluzívnu oblasť fyzikálnej paradigmy sa dostávajú do zovretia zmiešaných disciplín a metaforickej oblasti biologickej paradigmy. Biológia prinesie vo sfére spoločenských vied zásadnú premenu, podobnú tej, ktorú fyzika priniesla do oblasti organického.

Aristotelovská teória miestneho pohybu, ktorá sa pôvodne nachádzala v blízkosti eluzívnej oblasti, a teda sa matematizácii podľa Aristotelových vlastných slov vymykala,

bola v priebehu vedeckej revolúcie 17. storočia presunutá do oblasti novo sa rodiacej matematickej fyziky. Je veľmi pravdepodobné, že s nástupom biologickej revolúcie čaká rovnaké prehodnotenie na psychológiu, ekonómiu a sociológiu. Kognitívna psychológia, ekonómia a veľká časť sociológie budú asi z eluzívnej oblasti paradigmy novovekej vedy (z oblasti spoločenských vied) presunuté do centra novej paradigmy, ktorou je biológia. Súčasne s tým dôjde aj k zásadnej premene biológie. Podobne ako fyzika, ktorá vytvorila matematický opis miestneho pohybu, už nebola fyzikou v aristotelovskom zmysle; už ju necharakterizovala látka opisovaná pomocou štyroch aristotelovských príčin, ale stav opisovaný pomocou Newtonovej pohybovej rovnice. Je preto pravdepodobné, že aj biológia, ktorá absorbuje psychológiu, ekonómiu a veľkú časť sociológie, už nebude biológiou v tom zmysle, ako ju vymedzoval novovek. Nebude to teória živých systémov (teda disciplína definovaná v kontrapozícii k teórii neživých systémov, ktoré sú predmetom fyziky), ale bude to skôr teória systémov s biologickou informáciou (teda informáciou chápanou ako kód – v protiklade k teóriám informácie chápanej ako symbol). Z informačného hľadiska je kognitívna schéma, cena surovín či sociálna hierarchia veľmi podobná genetickému kódu. Informačný obsah vo všetkých týchto prípadoch môžeme vyložiť ako kód, reprezentujúci stupeň adaptácie systému na prostredie (problémovú situáciu, trh, sociálnu rolu či prírodné okolie). Na túto jednotu poukazuje rastúca úloha teórie hier v týchto disciplínach. Teda podobne ako disciplíny novovekej fyziky čerpajú svoju jednotu z pojmu stavu a z jeho opisu pomocou diferenciálnych rovníc, je pravdepodobné, že disciplíny modernej biológie budú čerpať svoju jednotu z pojmu informácie a jej opisu prostriedkami teórie hier. Pri tomto novom pohľade na biológiu sa psychológia, ekonómia a veľká časť sociológie stávajú biologickými disciplínami, nech už to znie akokoľvek čudne.



Keďže biologická revolúcia ešte nie je zavŕšená, náš nasledovný výklad je extrapoláciou doterajšieho vývinu. Zdá sa pravdepodobné, že *eluzívnou oblasťou biologickej paradigmy bude oblasť symbolického*. Treba však dodať, že detailné vymedzenie eluzívnej oblasti biologickej paradigmy, podobné Aristotelovej charakterizácii organického ako oblasti neprístupnej matematickému opisu a Descartovej charakteri-

zácie subjektívneho ako oblasti neprístupnej fyzikálnemu opisu, je otázkou budúcnosti. V tomto smere očakávame filozofickú reflexiu, ktorá vyloží protiklad pojmu informácie chápanej ako kódu (čo je pre biológiu paradigmatické chápanie pojmu informácie) a pojmu informácie chápanej ako symbol.[7] Rovnako možno predpokladať, že *metaforickú oblasť biologickej paradigmy budú tvoriť disciplíny príbuzné symbolickej antropológii*, teda disciplíny, ktoré síce pracujú s pojmom symbolickej informácie, ale snažia sa ju uchopiť na základe analýzy komunikácie v spoločnosti, čo je metafora prebratá z oblasti biológie (chápanej univerzálne ako vedy opisujúcej procesy tvorby, kopírovania a spracovania kódu).

Oblasť zmiešaných disciplín biologickej paradigmy budú pravdepodobne tvoriť disciplíny od sociobiológie, cez kognitívnu psychológiu až po kognitívnu antropológiu.[8] Ako určitú integrujúcu iniciatívu v tomto spletenci disciplín možno chápať kognitívne vedy. Ide o disciplíny, ktoré skúmajú rôzne aspekty subjektivity od vnímania, poznávania a pamäte cez jazyk a komunikáciu až po sociálnu hierarchizáciu a tvorbu identity. Postupujú však pritom ***pomocou experimentálnych metód***, ktoré boli vypracované v rôznych biologických disciplínach. Tento posun má za následok, že disciplíny, ktoré pôvodne patrili do eluzívnej oblasti fyzikálnej paradigmy (t. j. do sféry spoločenských

[7] Začína sa črtať konflikt v chápaní disciplín ležiacich na rozhraní metaforickej oblasti a oblasti zmiešaných disciplín biologickej paradigmy. Ako ilustráciu možno vziať Heglovou filozofiu, ktorá je podnes základom viacerých spoločenskovedných koncepcií. Zdá sa, že Heglov ***príbeh raba a pána*** z *Fenomenológie ducha* bude mať v rámci súčasnej vedeckej revolúcie podobný osud ako Aristotelova teória pohybu v priebehu revolúcie 17. storočia. Analogicky s Galileovou fyzikalizáciou teórie pohybu nebeských telies sa pokúsime načrtnúť biologizáciu príbehu raba a pána. Hegel prezentuje príbeh raba a pána ako emancipačný konflikt a otvára ho opisom boja, ktorý je údajne *bojom na život a na smrť*. Pozoruhodná na tomto boji na život a na smrť je však skutočnosť, že obaja zúčastnení prežijú. To je čudné, lebo podľa definície boj na život a na smrť nemá víťaza a porazeného. Končí sa jednou alebo viacerými mŕtvolami. To znamená, že boj, ktorý Hegel opisuje, nebol bojom na život a na smrť. Heglov pátos je len romantický opar, ktorý zakrýva čosi iné, čosi omnoho menej vznešené, menovite „*ruju*“. Heglov boj je štandardným bojom o samice, teritórium, zdroje a postavenie v sociálnej hierarchii, aký zvádza jeleň, pavián, tuleň či iný sociálny živočích každý rok. Cieľom Heglovej teórie bolo nahradiť kresťanský príbeh hriechu a spásy príbehom boja a emancipácie vďaka práci. Naším cieľom tu nie je vyjadrovať sa k detailom tohto príbehu, veď koniec koncov už bol nahradený radom iných, podobných emancipačných príbehov. Chceme iba ukázať, že sociobiológia otvára na celú túto problematiku zásadne nový pohľad. Pritom odpor, ktorý v čitateľovi tento „naturalizmus“ vyvoláva, je asi podobný odporu, ktorý pociťovali aristotelici, keď im Galileo spravil z Mesiaca obyčajný kus kameňa. Sociologické teórie konfliktu sú biologické teórie, a tak postupne prejdú do oblasti sociobiológie a uvoľnia priestor disciplínam, ktoré si všímajú symbolickú dimenziu ľudskej kultúry, teda dimenziu spadajúcu do eluzívnej oblasti biologickej paradigmy. Podobný osud postihne pravdepodobne aj teóriu odcudzenia. To, čo sa označuje týmto pojmom, je blízke tomu, čo prežíva každý jeleň, pavián či tuleň, keď ho silnejší samec odoženie od stáda a on stratí svoje postavenie v sociálnej hierarchii, ako aj prístup k samiciam.

[8] Tu sa zdanlivo dostávame do protirečenia s tým, čo sme o psychológii tvrdili, keď sme ju zaradili do metaforickej oblasti fyzikálnej paradigmy. Ale asociačná psychológia sa ako vedná disciplína zrodila v 19. storočí a vtedy spadala do metaforickej oblasti fyziky. V priebehu 20. storočia sa však predmet a metódy psychológie viackrát dosť zásadne zmenili. Mnohé výskumy, ktoré sa pôvodne robili výlučne na ľuďoch, boli prenesené aj na zvieratá. Evolučná psychológia, ktorá sa zrodila koncom 20. storočia, umožnila vysvetliť mnohé aspekty ľudského správania a prežívania na biologickej báze. Veľké časti psychológie sa tak presúvajú do sféry zmiešaných disciplín biologickej paradigmy.

vied), a ktorých metódy boli založené na introspekcii a empatii, dosahujú výsledky, ktoré zodpovedajú vedeckým štandardom biológie. Keď vymedzíme kognitívne vedy ako zmiešané disciplíny biologickej paradigmy a postavíme ich do jedného radu so zmiešanými disciplínami dvoch predošlých paradigiem (t.j. archimedovskou statikou a molekulárnou biológiou), tým len podčiarkneme skutočnosť, o ktorej sme hovorili vyššie. Práve na rozhraní zmiešaných disciplín a metaforickej oblasti biologickej paradigmy treba očakávať prelom smerom k novej paradigme, ktorá bude skutočnou paradigmou spoločenských vied, a teda umožní nastoliť ***normálny výskum v kuhnovskom zmysle***. Kognitívne vedy sú teda s veľkou pravdepodobnosťou cestou vedúcou k prvej spoločenskovednej paradigme.[9]

6 Stručné porovnanie s teóriou T. S. Kuhna

Predkladanú stať možno považovať za kritické vyrovnávanie sa s Kuhnovou teóriou vedeckých revolúcií. Náš odklon od Kuhnovej teórie začal uvedomením si skutočnosti, že existujú rôzne druhy vedeckých revolúcií. Ako príklad možno vziať newtonovskú a einsteinovskú revolúciu. Po newtonovskej revolúcii bola stará paradigma (aristotelovská fyzika) zavrhnutá a odstránená z vedeckej prípravy budúcich fyzikov. Naproti tomu po einsteinovskej revolúcii stará paradigma (newtonovská fyzika) nebola zavrhnutá a každý budúci fyzik absolvuje kurz klasickej mechaniky. Teda existuje zásadný rozdiel vo vzťahu medzi novou a starou paradigmou.

V stati [5] sme navrhli rozlíšiť tri druhy vedeckých revolúcií. Prvý druh vedeckých revolúcií sme nazvali *ideáciou* (jej príkladom je newtonovská revolúcia, pozri [9,10,12]). Ideácia fixuje základnú intenciu fyzikálnych teórií ako experimentálne určovanie faktov, ich matematickú redukciu na opis stavu, a hľadanie zákona opisujúceho časový vývin stavu. Druhý druh vedeckých revolúcií sme nazvali *re-prezentáciou* (jej príkladom je einsteinovská revolúcia završujúca zrod teórie poľa, alebo bohrovská revolúcia charakterizujúca vznik kvantovej mechaniky, pozri [11,13]. Tieto revolúcie priniesli zásadné zmeny v chápaní hmoty, času a priestoru, teda základných fyzikálnych kategórií. Ale ako v klasickej mechanike, tak aj v teórii poľa či kvantovej mechanike ostalo zachované, že fyzikálna teória za zakladá na opise stavu (ako udalosti v časopriestore či lúča v Hilbertovom priestore) a opise časového vývinu stavu (pomocou Maxwellových rovníc respektíve Schrödingerovej rovnice). Tretí druh vedeckých revolúcií sme nazvali *objektáciou* (jej príkladom je Lagrangeova a Hamiltonova revolúcia v klasickej mechanike, pozri [7]). Pri objektáciách sa mení opis stavu, inak sa volia primitívne pojmy a iným spôsobom sa opisuje interakcia. Ale ako newtonovská, tak aj lagrangeovská či

[9] Druhá nemenej zaujímavá zmena sa týka ekonómie. Je dosť pravdepodobné, že aj táto disciplína, alebo aspoň jej veľká časť, sa postupne presunie do sféry biológie. Postupne, tak, ako sa matematický aparát teórie hier, vyvinutý J. von Neumannom v 20-tych rokoch na opis ekonomického správania, presadzuje v evolučnej teórii, sa dá očakávať čoraz väčšie zbližovanie ekonomických a biologických disciplín. Matematický aparát teórie hier sa pravdepodobne stane štandardným aparátom biológie a bude mať podobnú zjednocujúcu funkciu, ako mal jazyk diferenciálnych rovníc pre fyzikálnu paradigmu. Pritom aj tento aspekt má paralelu vo vedeckej revolúcii 17. storočia, kde sa tiež prvý exaktný fyzikálny zákon, Galileov zákon voľného pádu, zrodil v oblasti, ktorá pôvodne spadala do aristotelovskej teórie miestneho pohybu. Pojem zrýchlenia pochádza teda z metaforickej oblasti antickej paradigmy - rovnako ako pojem stabilnej stratégie pochádza z metaforickej oblasti novovekej paradigmy (ekonómie).

hamiltonovská mechanika sú nerelativistické.

Odlíšenie troch druhov vedeckých revolúcií nás priviedlo k odmietnutiu Kuhnovej ***sociologickej*** teórie vedeckých revolúcií. Podľa nášho názoru pre jednotlivé druhy vedeckých revolúcií je možné podať racionálnu rekonštrukciu ich kognitívnej dynamiky. Problém je v tom, že Kuhn rôzne druhy vedeckých revolúcií nerozlišuje, a tak sa usiluje podať spoločný obraz všetkých revolúcií naraz. Keďže kognitívna dynamika je u každého z troch typov vedeckých revolúcií iná, zmiešaním všetkých troch druhov dohromady sa kognitívna dynamika stratí a čo ostane je len sociálna dynamika adaptácie vedeckej komunity na zmenu. Kuhnovu teóriu vedeckých revolúcií možno preto prirovnať k premietnutiu troch fotografií rôznych tvárí na jedno plátno. Premietnutím sa všetky individuálne črty vytratia, a čo ostane, sú neostré hrubé obrysy, spoločné všetkým trom tváram. Domnievame sa, že rozlíšením troch druhov vedeckých revolúcií je možné pokúsiť sa pre každý druh opísať jeho individuálnu kognitívnu dynamiku, a tak Kuhnovu sociologickú teóriu nahradiť teóriou epistemologickou. Prechod od sociologickej k epistemologickej rekonštrukcii umožňuje nahradiť Kuhnov *iracionálny* opis zmeny paradigmy ako zmeny geštaltu racionálnejším opisom založeným na teórii kognitívnej zmeny. Opísať kognitívnu dynamiku vedeckých revolúciu sa nám však podarilo len pre objektácie (ako zmeny formy jazyka – pozri [4,7]) a pre re-prezentácie (ako zmeny logickej, expresívnej, explanatorickej a integratívnej sily jazyka – pozri [6,11,13]). Napriek systematickému úsiliu sa nám nedarilo nájsť kognitívnu dynamiku pre ideácie. Preto v [8], keď sme porovnávali Kuhnovu a Lakatosovu rekonštrukciu dejín vedy, sme ešte zastávali názor, že Kuhnov iracionálny výklad revolúcie ako zmeny geštaltu síce neplatí pre objektácie a reprezentácie, ale pre prípad ideácií sme pripúšťali jeho správnosť. Vychádzali sme z predpokladu, že ideácia predstavuje tak zásadnú a tak hlbokú zmenu, že komunikácia medzi paradigmami tu nie je možná. Kuhnovu teóriu sme tak považovali za chybné zovšeobecnenie teórie platnej pre jeden druh vedeckých revolúcií na zvyšné dva druhy.

Medzičasom sme však dospeli k názoru, že Kuhnov iracionálny výklad vedeckej revolúcie ako zmeny geštaltu neplatí ani pre ideácie. Takto sme nakoniec dospeli k zásadnému odmietnutiu celej Kuhnovej teórie. Kuhn opisuje vedeckú revolúciu ako proces prebiehajúci *vovnútri paradigmatickej oblasti*. Podľa neho k vedeckej revolúcii dochádza vtedy, keď sa vo vede vytvorí kríza, teda keď sa nahromadí veľký počet anomálií, ktoré protirečia paradigme. Podľa nášho názoru v prípade vedeckej revolúcie v 17. storočí sa nová paradigma nezrodila v oblasti paradigmatických disciplín konštituovaných paradigmou antickej vedy. Práve naopak, nová paradigma sa zrodila na *rozhraní medzi oblasťou zmiešaných disciplín a metaforickou oblasťou*. To je dôležité, lebo sme presvedčení, že Kuhnov iracionálny obraz vedeckej revolúcie ako zmeny geštaltu je len dôsledkom toho, že dejiny vedy rekonštruuje pomocou zjednodušenej schémy, v ktorej chýbajú zmiešané disciplíny a metaforická oblasť paradigmy. Kuhn sa obmedzil výlučne na analýzu paradigmatických disciplín a celá iracionalita jeho obrazu vývinu vedy je artefaktom tohto zjednodušenia. Keď z obrazu vynechá prvky, ktoré sú nositeľmi kognitívnej dynamiky zrodu novovekej paradigmy, neostáva mu nič iné, než tento proces vyhlásiť za iracionálny. Zrod novej paradigmy v 17. storočí sa neudial vovnútri žiadnej paradigmatickej oblasti, ale došlo k nemu, keď sa astronomické objavy dostali do rozporu s aristotelovskou teóriou pohybu.

7 Úloha matematiky v spoločenských vedách

Otázka úlohy matematiky v rozvoji spoločenských vied je zložitá otázka. Nechcem ju tu diskutovať v jej plnej šírke, chcel by som len upozorniť na niekoľko analógií, ktoré sa ponúkajú v súvislosti s vyššie opísanou dynamikou vývinu vedy. Prvou je, že postoj, ktorý matematizáciu odmieta, ako napríklad metóda chápajúcej sociológie, nie je čosi zásadne nové, čo by nemalo v dejinách vedy obdobu. Chápajúca sociológia je v mnohom analogická s karteziánskou fyzikou (vysvetľujúcu tiaž pomocou víru jemnej látky), ktorú možno naopak považovať za „*chápajúcu*“ fyziku. Descartes sformuloval svoj systém bez matematiky a jeho kritériom vedeckosti je kritérium jasnosti a rozlíšenosti, teda vlastne kritérium jednoznačného chápania skutočnosti. Huyghens a Leibniz sa do konca života pridržali varianty tejto „*chápajúcej*“ fyziky a neboli ochotní akceptovať Newtonovu matematickú fyziku práve preto, že neumožňuje ***pochopiť***, ako vzniká príťažlivosť, ale túto len matematicky opisuje. Dokonca pytagorejská teória čísel, v ktorej sa číslam pripisovali vlastnosti ako spravodlivosť, možno označiť ako „*chápajúcu*“ matematiku.

Chápajúca sociológia či ekonómia je teda štandardná fáza, ktorou pri svojom vývine prejde každá vedecká disciplína. Dôležité je preto položiť si otázku, prečo je matematika „*matematická*“, teda prečo ľudia ako Thaitetos, Eudoxos a Euklides opustili pytagorejskú „*chápajúcu*“ matematiku a nahradili ju matematikou „*matematickou*“. Odpoveď je rovnaká ako odpoveď na otázku, prečo Newton, d‘Alembert a Euler opustili karteziánsku „*chápajúcu*“ fyziku a vsadili na fyziku „*matematickú*“. Dôvodom je to, že pytagorejská „*chápajúca*“ matematika rovnako ako karteziánska „*chápajúca*“ fyzika sú nekonzistentné. Hippasov objav nesúmerateľnosti a Newtonov objav trenia ukázali, že spomínané teórie obsahujú rozpor. Aj keď je prirodzené chcieť, aby sa svet čísel správal podľa našich intuícií o číslach, a aby príroda bola v zhode s našimi intuíciami mechanického pôsobenia, bohužiaľ to tak nie je. Akokoľvek nepredstaviteľná je nesúmerateľnosť strany a uhlopriečky štvorca a akokoľvek záhadné je newtonovské pôsobenie na diaľku, musíme ich akceptovať. Je to smutné, ale čísla ani príroda sa nesprávajú tak, ako by sme si želali, nedodržiavajú princípy „*chápajúcej*“ matematiky či fyziky. A „*matematická*“ matematika je zatiaľ jediný ***nástroj, ktorý umožňuje pracovať s kontraintuitívnym poznaním***.

Predovšetkým si teda treba uvedomiť, že matematika a fyzika sa nestali „*matematickými*“ z rozmaru. Stali sa takými z nutnosti, potom, ako vedci pochopili, že skutočnosť sa neriadi našimi intuíciami, presnejšie, že teórie vybudované na základe našich intuícií sú kontradiktorické. Podobne ako v matematike či vo fyzike aj v ekonómii či sociológii niet dôvodu, prečo by sa realita mala riadiť našimi intuíciami. Je pochopiteľné, že si ľudia želajú, aby sa zákonitosti ekonomiky riadili našim morálnym cítením, a aby dejiny nakoniec viedli k nastoleniu spravodlivého usporiadania spoločnosti. Ale takéto želanie nemá žiadnu vedeckú hodnotu. Mravné rozhorčovanie nám nepomáha porozumieť svetu. Čím skôr sa rozhodneme akceptovať skutočnosť takú, aká je (podobne ako sa Newton rozhodol prijať gravitačné pôsobenie a Eudoxos nesúmerateľné úsečky), tým väčšiu šancu máme porozumieť jej. Marxistická ekonómia podobne ako karteziánska fyzika a pytagorejská teória čísel sú po objavení sa „*matematickej*“ ekonómie, „*matematickej*“ fyziky a „*matematickej*“ matematiky iba pseudovedeckými špekuláciami bez akejkoľvek vedeckej hodnoty.

Druhou otázkou je, ako má vyzerať matematika, pomocou ktorej možno vybudovať ekonómiu ako exaktnú disciplínu. Prvé pokusy o vytvorenie matematickej ekonómie pochádzajú až od L. Walrasa a V. Pareta. Zdá sa, že pre ekonómiu vytvoril matematický aparát John von Neumann v podobe teórie hier. Kritici matematických metód v spoločenských vedách často nerozumejú zmyslu matematizácie a argumentujú proti nej zložitosťou predmetu, ktorá neumožňuje jednoznačný opis, ako ho poznáme z prírodných vied. Ale zmyslom matematizácie ***nie je*** v prvom rade poskytnúť bezpečné a jednoznačné poznanie. To je vedľajší produkt. Zmyslom matematizácie je v prvom rade odhaliť, čomu nerozumieme. Krásne to vidno na príklade Keplera a Galilea. Keplerova teória pohybu planét rovnako ako Galileova teória voľného pádu boli konceptuálne zásadne pomýlené. Dnes vieme, že planéty neunáša žiaden vír, ako si to predstavoval Kepler a rovnako vieme, že pri voľnom páde sa pohyb telesa nezrýchľuje sám od seba, ako si to predstavoval Galileo. Ale napriek tomu, že konceptuálne systémy, pomocou ktorých títo velikáni vedy 17. storočia uchopovali predmet svojho záujmu, boli nekonzistentné, vďaka tomu, že trvali na prísnej matematizácii svojich teórií, sa tieto nekonzistentnosti jasne manifestovali, a Newton tak mohol vytvoriť konzistentný opis mechanických javov. Preto úlohu matematických modelov v ekonómii či sociológii treba vidieť predovšetkým v demaskovaní nedôsledností a medzier v našom konceptuálnom porozumení predmetu týchto disciplín. Pokiaľ zotrvávame na čisto konceptuálnej úrovni, tak nám zložitosť predmetu a sugestívnosť našich princípov tieto nedôslednosti a medzery zakryje. Matematizácia nikdy nenahradí konceptuálne porozumenie. Umožňuje však voči nemu zaujať kritický odstup.

Poďakovanie: Ďakujem *Nadácii Alexandra von Humboldta* za udelenie štipendijného pobytu na Technickej univerzite v Berlíne, v rámci ktorého sa zrodila táto stať. Tento príspevok je súčasť grantového projektu VEGA číslo 1/3621/06 *Historické a filozofické aspekty exaktných disciplín*.

Literatúra

[1] Frank, Ph.: *Philosophy of science*. Englewood Cliffs, N. J.: Prentice-Hall 1957.
[2] Friederici, A.: Wie Sprache auf die Nerven geht. In: *Max Planck Forschung*, 2002/3, pp. 52-57.
[3] Kuhn, T. S.: *Štruktúra vedeckých revolúcií.* Bratislava: Pravda 1981.
[4] Kvasz, L.: History of Geometry and the Development of the Form of its Language. *Synthese* **116** (1998) 141-186.
[5] Kvasz, L.: On classification of scientific revolutions. *Journal for General Philosophy of Science* **30** (1999) 201-232.
[6] Kvasz, L.: Changes of Language in the Development of Mathematics. *Philosophia Mathematica* **8** (2000) 47-83.
[7] Kvasz, L.: Epistemologické aspekty dejín klasickej mechaniky. *Filozofia* **56** (2001) 679-702.
[8] Kvasz, L.: Lakatos' Methodology Between Logic and Dialectic. In: *Appraising Lakatos. Mathematics, Methodology and the Man*, 2002, eds. G. Kampis, L. Kvasz a M. Stöltzner, Kluwer, pp. 211-241.

[9] Kvasz, L.: Galilean physics in light of Husserlian phenomenology. *Philosophia Naturalis* **39** (2002) 209-233.
[10] Kvasz, L.: The Mathematisation of Nature and Cartesian Physics. *Philosophia Naturalis* **40** (2003) 157-182.
[11] Kvasz, L.: Epistemologické otázky fyziky: od antinómií čistého rozumu k expresívnym medziam jazyka. *Organon F* **11** (2004) 362-381.
[12] Kvasz, L.: The Mathematization of Nature and Newtonian Physics. *Philosophia Naturalis* **42** (2005) 183-211.
[13] Kvasz, L.: Epistemologické otázky modernej fyziky. *Organon F* **12** (2005) 40-61.

Abstrakce, komparace a generalizace jako základní operace lidského myšlení

Karel Pstružina[1]

Abstrakt. V příspěvku se budeme zabývat abstrakcí, komparaci a generalizací, které považujeme za základní operace lidského myšlení. Pokusíme se využít přístup k těmto operacím jak se objevuje v Kantově: Kritice čistého rozumu; případně v Heideggerově interpretaci této Kantovy práce; dále budeme vycházet ze srovnání mezi pojetím abstrakce a Husserlovými redukcemi. Příspěvek má vyústění v možnostech, které uvedené operace lidského myšlení poskytují např. při tvorbě pojmů.

1 Operace lidského myšlení

Analýza vědeckých metod stojí v popředí zájmu našeho výzkumu. Zdá se nám, že vědecké metody mohou poskytnout velmi stabilní prostředí pro identifikace pohybů lidského myšlení, neboť jejich výsledkem jsou stabilní a opakovatelné poznatky, k nímž se lidské myšlení propracovává vědeckým průzkumem jsoucen. Nejprve uvedeme výsledky pohybů lidského myšlení, které charakterizujeme jako operace lidského myšlení a dále se pak budeme zabývat speciálně třemi z nich, tedy abstrakcí, generalizací a komparací.

Operace lidského myšlení jsme seřadili od nejjednodušších k složitějším. Do první skupiny patří negativita a identita; do další skupiny již složitějších operací řadíme abstrakci, komparaci a generalizaci; a za nejsložitější považujeme symetrii, extrapolaci, eliminaci, kombinatoriku, trojčlenku a sylogismus. Domníváme se, že existují ještě další operace, která jsou však již natolik složité, že je můžeme považovat za vzorce a lidské myšlení je používá jako celky. Uvádíme tedy následující tabulku, v níž se snažíme identifikovat tyto operace lidského myšlení a zároveň se pokoušíme také říci, jaký pohyb při nich lidské myšlení vykonává. To nelze přiblížit jinak, než že tento pohyb přiblížíme prostřednictvím mluvy.

Tabulka 1. Operace lidského myšlení

Operace lidského myšlení:	**Popis:**
Negativita	*To není to*
Identita	*To je to*
Komparace	*To je jako to, nebo to není jako to*
Generalizace	*Tyto mají společné toto*
Abstrakce	*Toto charakterizuje to a toto necharakterizuje to; bez toho je toto stále tímtéž*
Symetrie	*To nastane proto, že takto je to v rovnováze k onomu*
Extrapolace	*Toto bude pokračovat takto, protože pokračování je funkcí onoho*

[1] Katedra filosofie, Vysoká škola ekonomická, W. Churchilla 4, Praha 3, 130 00, E-mail: Pstruzin@vse.cz

Eliminace	*Toto nepatří k oněm, protože se nevyznačuje tímto*
Kombinatorika	*Tyto lze zkombinovat do tolika těchto; toto lze doplnit tímto, abychom získali ono*
Trojčlenka	*Toto se má k tomu jako ono se bude mít k onomu*
Sylogismus	*Jestliže to se má k onomu jako ono se má k tamtomu, pak to se má stejně k tamtomu.*

Poznamenejme, že se nedomníváme, že tato tabulka operací lidského myšlení je úplná. Jsme si také plně vědomi toho, že nenahrazuje operace výrokové logiky, domníváme se však, že je vhodně doplňuje. Je zřejmé, že např. sylogismus lze zapsat prostřednictvím implikací a konjunkce, podle našeho názoru však lidské myšlení používá sylogismus již jako celek a nemusí neustále probíhat po cestách implikací a konjunkce, chce-li zjistit zda Sokrates je smrtelný, jestliže člověk je smrtelný a jestliže Sokrates je člověk. Obdobně je tomu i u všech dalších operací lidského myšlení. Zaměřme se nyní na výše uvedené tři operace, tzn. na abstrakci, komparaci a generalizaci a pokusme se jejich prostřednictvím prokázat jaké pohyby lidské myšlení při nich vykonává.

2 Kant a jeho pojetí pojmotvorby

Nejprve se chceme zamyslet nad tvorbou pojmu, a jak tento problém řeší I. Kant.

Čteme-li Kanta [4], pak se od něj dovídáme především to, že pojem je obecné v jednotlivém (nebo konkrétním). A dále, že tvorba pojmu se zakládá na třech operacích: reflexi, komparaci a abstrakci, přičemž nejdůležitější je reflexe sdružující a zahrnující pod sebe další dvě výše uvedené operace. Tyto tři operace nejsou vzájemně nezávislé, nebo snad po sobě následující, ale jde o jedinou složitou operaci, při níž se vyhledává něco, co je obecné v jednotlivině a zároveň se toto srovnává s dalšími jednotlivinami, zda se v nich tato obecnost nachází.

Dále poznamenáváme také to, že Kant, uvažuje-li o abstrakci jako jedné z operací angažovaných při tvorbě pojmu, přiklání se k abstrakci jako abstrakci od něčeho, nikoli jako abstrakci něčeho. Heidegger v této souvislosti poznamenává: *Abstrakce tedy pro Kanta není vyhlédnutím jednoty, ale odhlédnutím od různosti* [2 str. 196].

Jak tedy pojem podle Kanta vzniká. Kant klade důraz na reflexi, což je operace lidského myšlení, která je schopna zrcadlit něco (co již bylo vyhlédnuto a co tvoří obecné v jednotlivé konkrétní věci) ve vědomí, při čemž je právě v této reflexi něco *obejmuto* a tak *pojmuto* a *přijmuto* pod pojem. To, co je takto přijmuto (zajato) je obecnost, ale zároveň obecnost v rozmanitosti. Při reflexi jde tedy o spontánnost vyhlédnutí (přihlédnutí) něčeho obecného, což pak dále srovnáváme s dalšími konkrétními věcmi a tak rozlišujeme to, co je pro každou konkrétnost specifické a zároveň abychom zjistili, co tyto konkrétnosti mají společné.

Zde jsme u problému, kde se ve svém pojetí odlišujeme při objasnění vzniku pojmu od Kanta a proto musíme své stanovisko hlouběji objasnit.

3 Abstrakce

Naše stanovisko je tedy následující.

Při vzniku pojmu jde o souhru tří operací: komparace, generalizace a abstrakce, které na sebe navazují a vzájemně se doplňují. Nejde zde jen o nahrazení Kantovy reflexe generalizací, ale především o to, že lidské myšlení je vždy rozpolceno ve dva proudy (primární a sekundární proud lidského myšlení), z nichž druhý automatický reflektuje, tzn. ohýbá a zrcadlí vše, co je primárním proudem vykonáváno a takto je to přiváděno k uvědomění. Východiskem pojmotvorby je podle našeho názoru nalezení podstaty (tedy nikoli jen obecniny), která je komparována s podstatou v jiných věcech a jevech a teprve tak je nalézáno něco obecného v jednotlivém.

K podstatě se dostáváme prostřednictví abstrakce, avšak abstrakce má podle našeho názoru jinou podobu, než je o ní často uvažováno. Nejde o odtahování (odmršťování), abstrakce není aktem, kdy od věci jako celku dospíváme k jednotlivým jejím singularitám, tedy charakteristikám, k vlastnostem či rozmanitostem, ale ty jsou v mysli již předem.

Zde se také odlišujeme od Husserla a jeho fenomenologické redukce. Podle Husserla jde při fenomenologické redukci o uzávorkování jakéhokoli zaujímajícího stanoviska k objektivnímu světu. Co to však znamená, že uzávorkujeme objektivní svět? Takovéto pojetí není ničím jiným, než odtažením něčeho nepodstatného a odmrštěním, takže nám stále něco zůstává, co dělá věc věcí. Jde tedy o abstrakci. Avšak abstrakce pro Husserla není ochuzením, ale naopak obohacením, protože při ní sice redukujeme (tzn., že uzávorkujeme individuální jestvování objektu), ale zároveň produkujeme něco, co není ve světě; a tím co produkujeme jsou podstaty, čisté ideality, odtržené od všeho individuálně zkušenostního. Vraťme se však k redukci. Jak však můžeme jen tušit, co máme odmrštit? Napadají nás dvě možnosti:

- Buď si budeme počínat tak, že to zkusíme a v obrazotvornosti mysli si budeme představovat, zda se věc může obejít bez onoho odmrštěného a tudíž toto odmrštěné nepatří k eidos (podstatě věci);
- nebo musíme mít nějakou předtuchu toho, co je podstatou a na tomto základě určíme i nepodstatnosti.

Husserl odpovídá, že tím, že provedeme tuto epoché (že uzávorkujeme svět) získáme sebe sama *jakožto já, a to s vlastním čistým životem vědomí, v němž a skrze nějž je veškerý objektivní svět pro mne – a to způsobem jaký je právě pro mne. … Svět pro mne není vůbec nic jiného než svět, který je v takovém cogito vědomě jsoucí a mající pro mne platnost* [3 str. 24]. Pro Husserla tedy máme v sobě celý svět. *Tak tedy přirozenému bytí světa – toho světa, o němž mluvím a mohu mluvit – předchází de facto bytí čistého ego a jeho cogitationes jakožto bytí o sobě dřívější* [3 str. 25].

Náš přístup je v něčem jiný. Domníváme se, že jakákoli konstituce věci (a mentálu) jako celku je totiž vždy skladbou z jednotlivých vlastností, ze stimulů, které byly senzory vytrhávané ze jsoucen, a které jsou zde již předem jako singulární. Jsou tím, co lidské myšlení rozpoznává prostřednictvím operací identity a negativity, a tyto vlastnosti (rozmanitosti) jsou lidským myšlení slučovány (dávány dohromady) tak, že si uvědomujeme konstituovanou věc (např. hrníček) jako celek. Vlastnosti mají svůj základ ve jsoucnech určených myšlením a jsou lidským myšlením jen rozpoznány jako

specifikum tím, že jsou přiřazeny k něčemu, co již bylo vymezeno pohybem identity a negativity v enkapsické struktuře endoceptů, nebo co musí být teprve vymezeno strukturními vztahy k již vymezenému. Vlastnost je tím, co se utváří a co vyniká ve vztazích. Bez toho, aby něco bylo ve vztahu, nemůže vyjevit, čím vlastně je, nemůže vyjevit svou podstatu.

Jednotlivé stimuly tedy přichází do mozku dokonce odlišnými neuronálními dráhami a jsou také v odlišných částech mozku identifikovány. Např. zrakové centrum mozku je rozděleno do několika oblastí, v jedné z nich se např. identifikují vertikály a v jiné oblasti zase ty ze stimulů, které vyznačují horizontály, v další části mozku se odlišují barvy, v jiné je identifikován pohyb, atd.

V současnosti je u makaka známo 32 zrakových korových oblastí propojených 305 vzájemnými spoji, z toho 7 oblastí zpracovává kromě zrakových podnětů i podněty další. … Neurony každé z oblastí se chovají odlišně a jednotlivé oblasti mají i odlišné vstupy a výstupy [1 str. 154].

Takovéto singulární stimuly jsou pak lidským myšlením syntetizovány[2] do konečné podoby, což zde nazýváme konstitucí věcí a mentálů.

My nabýváme dojmu, že jsme provedli abstrakci, že jsme odmrštili některé vlastnosti, které neurčují podstatu, ale vlastně jsme se jen dostali k singulárním stimulům tak, jak je lidské myšlení vytrhlo ze jsoucen. To, čemu říkáme abstrakce je pro nás orientace perpetuálního pohybu lidského myšlení v před-konstitutivním stádiu, v němž jsou identifikovány jednotlivé vlastnosti, nebo charakteristiky věci. Posobnost projikcí lidského myšlení (jejich následnost prožívaná jako čas) však může někdy narazit na další vlastnosti, další charakteristiky, které musí být identifikovány a zahrnuty pod nový pojem, tzn. že je nutné vymezit je jako specifické a zařadit jako takové do vztahů vůči něčemu již vymezenému. Vlastnost sama o sobě (tzn. jakožto jednotlivé stimuly vytržené ze jsoucen) není podstatou.

Pohybuje-li se lidské myšlení v předkonstitutivním stádiu, může současně docházet ke komparacím, tzn. k takovým pohybům lidského myšlení, v němž jsou aktuálně iniciované neuronální okruhy vzájemně srovnávány s neuronálními okruhy (paměťovými stopami) dříve ze jsoucen vytržených obdobných stimulů. Vše aktuální je podrženo a nastavováno k reflexivním reverberacím.[3] Tzn. že primární proud lidského myšlení podržuje aktuální stimul a druhý z proudů probíhá enkapsickou strukturou a hledá obdoby. Tím dochází k identifikaci všeho, co je pro konstituovanou věc shodné s předchozími, v mysli uchovávanými vjemy. Z těchto shodně reverberovaných neuronálních okruhů jsou eliminovány ty, které při některé z předchozích projikcí chyběly a přesto lidské myšlení z těchto stimulů konstituovalo obdobnou věc, a takto je hromadně komparováno vše zařazené pod stejnost (pod stejný pojem), až se jasně rozliší to, co tvoří eidos, tedy to, čím věc je, co je její podstatou, bez čeho již věc není tou určitou věcí. Tyto procesy jsou spontánními a do značné míry neuvědomovanými. Jsou to procesy, které zná již dítě, ale které neumíme definovat a kterým se téměř neumíme naučit. To znamená, že jde o automatismy, který mohou být kantovsky apriorními procesy.

---

[2] M. Heidegger v této souvislosti, tedy v souvisosti sdružování smyslových stimulů, počitků, hovoří o syndóze. I. Kant však používá termín synopse.

[3] Předpokládáme, že takovýchto aktuálních stimulů je primární proud schopen najednou podržet 7 ± 2 a komparovat je se stejným počtem, který nastavuje sekundární proud. Tato naše domněnka se opírá o proslavenou stať: Miller, G.A.: The Magical Number Seven, plus or minus Two [5]

Abstrakcí jsou tedy až tyto eliminativní operace lidského myšlení v předkonstitutivním stádiu. Avšak podstatu si nemůžeme představovat tak, že jde jen o jedinou charakteristickou vlastnost (reprezentovanou jediným neuronálním okruhem). Mnohdy může jít o souhrn několika takovýchto jednotlivých vlastností, které teprve tím, že jsou ve vzájemných vazbách dávají podstatu a jsou obecninou v jednotlivém.

4 Generalizace a komparace

Při vzniku pojmu tedy musíme vzít v úvahu ještě také to, že abstrakce je současně doplňována generalizací, že tedy lidské myšlení nejen vylučuje vše, bez čehož je něco stále sebou, ale že se také při neustálé aktivaci a podržení podstaty primárním proudem lidského myšlení sekundární proud pohybuje tak, že to, co je identifikováno jako podstata srovnává s jinými konkrétními věcmi a přihlíží k tomu, zda jsou tyto charakteristiky (tato podstata) také v nich přítomny. Znamená to tedy, že opět na úrovni jednotlivostí dochází k jiným reverberacím, a sice těm, které signalizují shodnost podstat, tedy obecnost. To je pak pojmenováno a zajmuto pod pojem. Aby něco bylo pojmenováno musí být aktivováno také jazykové centrum mysli s vhodnými adekváciemi, tzn. výrazy, které tuto podstatu vystihnou. Tato aktivita jazyka nám dále napovídá o jaké pohyby lidského myšlení se jedná.

Generalizace však také zjišťuje rozmanitost konkrétností a tím i specifika, či případně individuální odchylky od podstaty, které však nebrání pojmenování, tedy obejmutí a zahrnutí (zajmutí) pod pojem. Rozmanitost pak navíc posluhuje zařazení věci do intencionální struktury prostřednictvím asociativních vazeb, jde tedy o jeden z příznaků komplexity při vymezení intence [6 str. 159].

Pojem je tedy zobecněním podstaty a vymezením podstaty jako něčeho obecného v konkrétnosti, k čemuž jsme došli abstrakcí, generalizací a komparací při kooperaci primárního a sekundárního proudu lidského myšlení. Jde tedy o sjednocení, tedy jak vyhlédnutím jednoty, tak i odhlédnutím od růzností, neboť vyhlédnutí něčeho podstatného, co je zobecnitelné, je zároveň i odhlédnutím od toho, co je různé, ale nezobecnitelné. Jak jinak bychom mohli k vyhlédnutí jednoty dospívat?

Toto vše se však odehrává jen na úrovni konstituce věcí, tedy smyslovosti a afekcí. Jak tomu však je, když tvoříme pojem, který nemá jako svůj protipól věc (resp. stimuly vytržené ze jsoucen), ale kdy jde jen o mentály, nebo o vystižení a rozpoznání vztahů mezi jsoucny?

5 Čisté rozvažovací pojmy

V takovýchto případech může jít o obdobný postup, jenže lidské myšlení zde operuje nikoli s věcí, ale s mentálem, s představou. Podstata, obecnina v jednotlivém je však v těchto případech obtížně zachytitelná, neboť představa se může v průběhu lidského myšlení proměňovat a nastupovat jako něco pozměněného; nebo lidské myšlení neudrží soustřednost na představě a uchyluje se do asociativních zřetězení svého pohybu a tak se vzdaluje možnosti prostřednictvím abstrakce, komparace a generalizace postihnout to, co je obecné v představách.

Tak, jako když máme před sebou hrníček, může se lidské myšlení soustředit na představu a nejen hrníčku, ale např. toho, co je zrada, či láska (či cokoli jiného). Jsme tedy v situaci, kdy se zrada nebo láska nějak objevuje jako určité prožívané a prociťované

vztahy (obdobně tomu, jak jsme ze singularit konstituovali hrníček). Následně tedy lidské myšlení tyto prožitky začne oprošťovat od jednotlivých charakteristik toho, co pociťujeme, jsme-li zamilováni, nebo máme-li pocit zrady. Postupně tak zjišťujeme, že láska je pocitem, jsou-li současně pociťovány takové a takové prožitky (např. prožitky vzrušení, krásy, porozumění, atd.). Takto se postupně dostává lidskému myšlení možnost, aby eliminovalo další podružné doprovodné jevy a dostávalo se k podstatě kterou pojmenuje. Lidské myšlení tedy může i ve složitých vztazích a vlastnostech začít diferencovat, rozlišovat, nebo rozdělovat.

Zatímco eliminace nepodstatnosti se u věcí dělá na úrovni již rozloženého, tzn. proto, že senzory vytržené stimuly jsou v singulární formě a lidské myšlení je musí nejprve syntetizovat do podoby věcí tím, že věc konstituuje na jsoucnu, tak při pojmotvorbě čistých rozvažovacích pojmů (Kant v této souvislosti píše o takových uzurpovacích, nebo rozvažovacích pojmech jako je štěstí nebo osud [4 str. 100][4]), je zde prvotně prožívaná situace a z ní vyplývající mentál jako pociťovaný celek. Chceme-li vystihnout jeho podstatu, musí lidské myšlení tento celek nejprve nějak rozdělit, aby v něm mohlo rozlišit různé aspekty. Takto rozdělené a rozlišené aspekty jsou pak komparovány s minulostními prožitky, tedy s paměťovými stopami jako složkami zkušeností (s retencemi) a je hledáno obecné v jednotlivých prožitcích, čímž vyhledáváme to, zda takováto situace aktuálního prožitku je podobná již něčemu, co jsme prožili. Lidské myšlení se tedy pohybuje uvnitř mysli a spojuje jednotlivé retence, vyhledává je, nebo něco podobného v endoceptivní struktuře a tvoří pojem. Tento případ je tím (i když odlišně od našeho příkladu), kdy se Kant domnívá, že v pojmotvorbě je zapojena obrazotvornost.

S takovýmto postupem by však Kant nesouhlasil, neboť podle něj je možné čisté rozvažovací pojmy tvořit čistou dedukcí. Musí však jít o transcendentální a nikoli empirickou dedukci (což je výše uvedený případ). Co to znamená? Především snad to, že uplatňujeme postup od obecného k jedinečnému (neboť takto Kant převážně dedukci chápe a nikoli jako jakoukoli inferenci) jen v mysli s představami, aniž bychom to srovnávali s něčím, co by po způsobu Heideggera bylo možno nazvat průměrnou zkušeností, která se, jak se domníváme opírá o prožitky.

Zdá se nám, že v takovémto případě proměřujeme situaci s asociativními vazbami, které jsou situací generovány a již jsou pojmenovány a tím dochází k vymezení významu pojmu skrze sémantickou síť. Vymezení čistého rozvažovacího pojmu sémantickou sítí je však už naší interpretací, než přesným zachycením toho, jak je pojem skrze transcendentální dedukci vymezován Kantem.

Kantovo deduktivní vymezování čistých rozvažovacích pojmů má své vyústění v kategoriích, které jsou pojmy o předmětu vůbec a jejichž prostřednictvím je předmět určen [4 str. 106-107]. Uvažuje tedy o kategoriích jako o něčem, co se vztahuje k předmětům, ale nikoli jako k nějakým konkrétním předmětům, ale k tomu, jak vůbec mohou být předměty určeny vzhledem k některé z logických funkcí souzení, tedy jak mohou být myšleny. Dále: kategorie se vždy musí vztahovat ke smyslovosti jako podmínce vzniku názoru, i když vznikají nezávisle na smyslovosti, ale pouze v rozvažování [4 str. 114]. Jde tedy o vymezení jejich úlohy při myšlení a poznávání.

[4] Podotýkáme, že Heidegger [2 str. 251] se snaží o určení „objektivní reality těchto pojmů z *průměrné zkušenosti o sobě, kterou prochází lidský pobyt*“.

Kategorie jsou tedy podmínkami myšlení, právě tak, jako prostor a čas jsou podmínkami názoru [4 str. 527] a v kategoriích je počátek transcendentální dedukce, který vyústí v čistých rozvažovacích pojmech.

Pokud se snažíme zjistit jaké jsou pohyby lidského myšlení, pak můžeme využít tyto Kantovy úvahy následujícím způsobem.

Podle našeho názoru jsou kategorie něčím, co umožňuje strukturní pohyb lidského myšlení mezi již konstituovanými mentály. Lidskému myšlení jsou kategorie vlastní. Jde tedy o to, že lidské myšlení s konstitucemi věcí a mentálů zároveň konstituuje něco, co je pro věc a vjem (mentál) důležité a tak je seskupuje, katalogizuje. Takovými důležitými určeními mohou být např. místo, čas, trpnost, nebo činnost, dále to, zda jsou nutným důsledkem, či výjimečným, specifickým, zda konstituují to, co je podstatou nebo co je jen jevícím se atributem, co, je obecninou, nebo specifičnosti jednotlivé věci nebo jevů, atd. Když pak lidské myšlení něco dalšího konstituuje, rozpoznává vše nejprve z hlediska těchto a dalších kategorií a proto jeho pohyb již nemusí probíhat vším již konstituovaným, ale jen tím, co je již setříděno, např. podle podstaty, nebo podle místa výskytovosti, nebo podle trpnosti, či činnosti, atd. Takže kategorie jsou tím, co je na věcech a vjemech tím nejobecnějším a co se netýká konkrétních věcí a vjemů. Domníváme se však že kategorie nemusí vznikat jen z čistého rozvažování a být tedy předchůdným určením pohybu lidského myšlení, ale že kategorie mohou vznikat obdobně jako pojmy i ze zkušenostních rovin konstitucí věcí a vjemů, i když nebudou nikdy reprezentací věcí, ale jen idealit.

Poděkování: Tento příspěvek vznikl za podpory grantové agentury GAČR 401/05/0015 v rámci grantového úkolu Vzorce operací lidského myšlení.

Literatura

[1] Crick, F.: *Věda hledá duši*. Praha: Mladá Fronta 1997.
[2] Heidegger, M.: *Fenomenologická interpretace Kantovy Kritiky čistého rozumu*. Praha: Oikúmené 2004.
[3] Husserl, E.: *Karteziánské meditace*. Praha: Svoboda 1968.
[4] Kant, I.: *Kritika čistého rozumu*. Praha: Oikúmené 2001.
[5] Miller, G.A.: *The Magical Number Seven, plus or minus Two: Some Limits on our Capacity for Processing Information*. *Psychological Review* **63**/2 (1956) 81-97.
[6] *Pojednání o lidském myšlení*. Praha: Ekopress 2005.

Mozgové štruktúry a mechanizmy riadenia strategickej pozornosti

Rudolf Andoga, Norbert Kopčo[1]

Abstrakt. V každodennom prostredí sú zmysly človeka nepretržite vystavené množstvu zložitých, súčasne prichádzajúcich podnetov. Strategická pozornosť nám umožňuje efektívne sa orientovať v takomto prostredí, a to optimálnym výberom behaviorálne dôležitých podnetov a sústredením limitovanej kognitívnej kapacity mozgu na ich spracovanie. Cieľom tohto príspevku je poskytnúť prehľad o neurálnych štruktúrach podieľajúcich sa na riadení strategickej pozornosti s dôrazom na vizuálne, sluchové, a krosmodálne mechanizmy riadenia pozornosti. Článok taktiež sumarizuje existujúce modely strategickej pozornosti, zdôrazňuje otvorené otázky a navrhuje alternatívne hypotézy, ktoré môžu byť východiskom pre budúce experimentálne a teoretické štúdie riadenia strategickej pozornosti.

1 Úvod

Vnímanie a reprezentácia priestoru sú dôležitými kognitívnymi schopnosťami živých a umelých organizmov [5], nevyhnutnými pre plynulú orientáciu, pohyb a interakciu organizmu s okolím. Človek používa na orientáciu v priestore primárne zrak. Ale k priestorovému vnímaniu prispievajú aj ostatné zmysly (sluch, hmat, atď.) a spôsob, akým si vyberáme, ktorej modalite dáme prednosť, alebo ako informácie z rôznych modalít kombinujeme, je predmetom intenzívneho výskumu.

Vnímanie a správanie sa živočíchov a človeka nezáleží len na priamej stimulácii jednotlivých senzorických orgánov, ale aj na tom, ktorým aspektom tejto stimulácie v tom ktorom momente venujeme pozornosť [23]. Pojem „selektívna pozornosť“ popisuje procesy, ktoré nám umožňujú selektívne spracovávať niektoré z prichádzajúcich senzorických podnetov tak, aby sa informácie dôležité pre našu aktivitu, alebo podnety s veľkým biologickým významom, spracovali detailnejšie než ostatné prichádzajúce podnety.

Tradične sa výskum selektívnej pozornosti zameriaval zvlášť na jednotlivé senzorické modality, počnúc klasickými experimentmi študujúcimi selektívne počúvanie [2,14] a pokračujúc štúdiom vizuálneho vnímania (napr. [18]). V reálnych prostrediach je ale veľká väčšina podnetov vnímaná multimodálne. Napríklad aj v klasickom príklade sluchovej selektívnej pozornosti – v tzv. koktail párty situácii, keď sa snažíme počúvať jednu z niekoľkých súbežných konverzácii [14] máme k dispozícii nielen sluchové podnety, ale aj vizuálne podnety (napr. pohyb perí, výraz tváre a gestá hovoriaceho, na ktorého sa sústredíme). Naviac, informačná hodnota jednotlivých modalít sa môže dynamicky meniť, napr. keď sa zvýši hlučnosť v miestnosti (kedy by sme sa mali viac sústrediť na vizuálne podnety) alebo keď sa tvár hovoriaceho na chvíľu stratí (kedy sa zas musíme sústrediť na to, čo počujeme). Napriek tomu ale prevládajúci model selektívnej pozornosti sedemdesiatych a osemdesiatych rokov 20. storočia uvažoval o každom vnemovom systéme ako o samostatnom výpočtovom prostriedku s obmedzenou kapacitou, v ktorom pozornosť slúži na riadenie prideľovania výpočtovej kapacity pre spraco-

[1] Katedra kybernetiky a umelej inteligencie, FEI TU Košice, Email: andoga@neuron.tuke.sk, kopco@bu.edu

vanie najdôležitejšieho z prichádzajúcich podnetov. Čiže pozornosť je špecifická pre danú vnemovú modalitu a nezávislá od ostatných modalít (napr. [22,24]).

Že sa o pozornosti nedá rozmýšľať ako o súbore nezávislých mechanizmov pracujúcich oddelene pre každú vnemovú doménu ukázali ako prví Driver a Spence [8]. Vo svojom experimente prezentovali súbežne cez dva monitory dva rôzne záznamy rozprávajúceho človeka, pričom zvuky boli prezentované v jednom meraní z reproduktorov zodpovedajúcich obrazu a v druhom meraní z navzájom vymenených reproduktorov. Úlohou subjektu bolo stále počúvať čo sa hovorí na jednom z monitorov, pričom o experimentálnej manipulácii so zámenou monitorov bol subjekt informovaný. Takže jeho úlohou v prvom meraní bolo sústrediť svoju vizuálnu aj sluchovú pozornosť na to isté miesto, zatiaľ čo v druhom meraní potreboval sústrediť vizuálnu pozornosť na jeden monitor a sluchovú pozornosť na polohu reproduktoru druhého monitora. Presnosť identifikovaných slov bola v druhom meraní signifikantne nižšia než v prvom meraní, čo značí, že schopnosť človeka sústrediť svoju vizuálnu pozornosť na jedno miesto a sluchovú pozornosť na iné miesto je obmedzená, a že mechanizmy strategického riadenia zrakovej a sluchovej pozornosti nie sú navzájom nezávislé.

Nasledujúce kapitoly podávajú prehľad o súčasnom stave poznatkov o priestorovom riadení endogénnej (strategickej, vedomím riadenej) selektívnej pozornosti, a o jej význame pre integráciu a interakcie medzi jednotlivými zmyslovými vnemami.

2 Rozdelenie pozornostných mechanizmov

Jednou z hlavných téz o pozornosti je, že pozornosť nie je jedna entita, ale súbor procesov, ktoré prebiehajú v neurálnom systéme. Tieto procesy môžu navzájom interagovať s inými procesmi v mozgu, a podieľať sa tak na rôznych percepčných, kognitívnych a motorických úlohách. Pozornosť akéhokoľvek podsystému neurálneho systému nie je možné jednotne definovať. Je možné sa na ňu pozerať z hľadiska procesov redukcie informácií na vstupe systému, alebo ako na selektívny proces pre prácu s cieľovým objektom v prostredí s distraktormi. Je možné ju chápať ako proces vyhľadávania a v konečnom dôsledku aj selektívny klasifikačný proces, alebo ako proces tvorby konceptov (vizuálnych, senzomotorických). Pozornosť je možné rozdeliť do dvoch základných skupín a to na:

- **automatickú** pozornosť, vyvolanú podnetom z vonkajšieho prostredia, tiež nazývanú exogénnou alebo senzorickou pozornosťou. Príkladom fungovania automatickej pozornosti je ilúzia vykresľovanej čiary [20]
- **strategickú** (vedomím riadenú, dobrovoľnú, endogénnu) pozornosť, ktorá charakterizuje proces, keď sa človek vedome rozhoduje, či sa sústredí na vonkajšie podnety, vnútorné myšlienky, alebo na iné kognitívne procesy a aktivity.

Toto rozdelenie, akceptované aj dnešnými teóriami, prezentoval už James v roku 1890. Ďalšie a všeobecnejšie rozdelenie pozornostných procesov môže byť v zmysle nasledovných troch komponentov: *selekcie, sledovania* a *riadenia*. Selekcia je zrejme zásadný a základný proces pozornosti, pričom existujú tri hlavné hypotézy o jej fungovaní. Podľa nich je selekcia založená buď na *umiestnení objektu – location based* [3], na *objekte samotnom*, t.j. že je *objektovo orientovaná – object oriented*, [10], alebo že je *objektovo znakovo založená – object token based* [15]. *Sledovanie* alebo udržiavaná pozornosť je dôležitým aspektom pri cielenej interakcii s objektom v prostredí, v ktorom sa nachádzajú

aj rušivé objekty (distraktory). Sledovanie zaručuje, že ciele sú v čase udržiavané. *Riadenie* pozornosti je proces, pri ktorom je systém schopný zmeniť stálu pozornosť z jedného objektu na iný objekt, ktorý sa stáva novým cieľom interakcie.

Z hľadiska modelovania priestorového vnímania je dôležité poznať kedy (za akých okolností) je pozornosť založená na polohe objektu a za akých na objekte samotnom. Taktiež je základnou otázkou ako sa zameranie priestorovej pozornosti riadi.

3 Modality pozornostných procesov

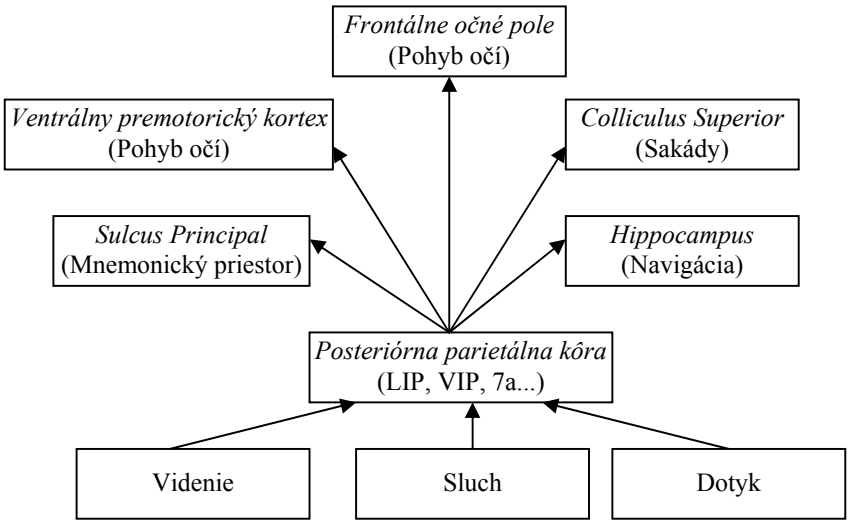
Množstvo každodenne vnímaných objektov a udalostí generuje pozornostné vodítka (anglicky cues), ktoré sa objavujú buď v jednej modalite, alebo nejakým stupňom pokrývajú viacero zmyslových modalít. Kapacita mozgu určená pre koordináciu rôznych senzorických signálov prichádzajúcich z jedného alebo viacerých zdrojov je základná vlastnosť, ktorá nám vo výsledku poskytuje jednotný vnem okolitého sveta. Táto koordinácia je tiež základom pre smerovanie (riadenie) pozornosti a riadenie pohybu v rámci nej. Predošlým výskumom bolo jasne preukázané, že neurálna syntéza multisenzorických vnemov môže zlepšiť pravdepodobnosť detekcie a odpovede na nejakú udalosť, ale tiež aj schopnosť identifikovať a lokalizovať danú udalosť. Na druhej strane, ak je rôznym modalitám prezentovaná konfliktná senzorická informácia, pravdepodobnosť správnej lokalizácie a identifikácie sledovanej udalosti sa môže znížiť. Na ilustráciu takejto situácie môže poslúžiť známy McGurkov efekt, ktorý ukazuje, že aj keď porozumenie počutej reči sa môže čítaním z pier značne zlepšiť, počúvanie človeka vyslovujúceho jednu hlásku a dívanie sa na toho istého človeka pri vyslovovaní inej hlásky vedie k vnemu, ktorý nezodpovedá ani videnému ani počutému podnetu.

Aj keď náš primárny záujem sa sústreďuje na fungovanie pozornosti pri vizuálnom a sluchovom priestorovom vnímaní, pre pochopenie týchto interakcií je dôležité aj skúmanie interakcií s dotykovými vnemami. Neurálne systémy zodpovedné za základné spracovanie vnemov jednotlivých zmyslových modalít sú v mozgu priestorovo i funkčne oddelené. Pre behaviorálne skúmanie ich fungovania a interakcie medzi nimi sa používajú psychofyzikálne merania na ľudských subjektoch [15].

4 Neurálna báza krosmodálnych pozornostných vplyvov

Pre identifikáciu fyziologických závislostí procesov multimodálnej pozornosti sú používané elektrofyziologické štúdie na zvieratách. Obrázok 1 ukazuje závislosti jednotlivých modalít a neuroanatomické oblasti zodpovedné za ich spracovanie.

Oblasťou mozgu zodpovednou za spracovanie vnemov rôznych modalít je posteriórny parietálny kortex (PPC), ktorý predstavuje rozhranie medzi vnemami a následnou akciou [7,19,21]. Experimentálne boli u neľudských primátov zistené multimodálne oblasti PPC a to ventrálny interparietálny kortex (VIP), laterálny intraparietálny kortex (LIP), polysenzorická zóna (PZ). Neuróny v týchto oblastiach odpovedajú na stimuly rôznych modalít. Relatívne dobré je súčasné porozumenie krosmodálnym interakciám v oblastiach PZ a VIP.



Obrázok 1. Schematická reprezentácia štandardnej teórie multisenzorickej integrácie (podľa Grossa a Graziana 1995)

Neuróny oblasti PZ sú citlivé hlavne na dotykové a vizuálne vnemy z blízkeho priestoru hlavy a hornej časti tela. Pre 95% buniek oblasti PZ platí, že ich receptívne pole pokrýva iba časti tela kontralaterálne k hemisfére, ktorú skúmame. Pre temer všetky bunky platí, že vizuálne receptívne pole je obmedzené hĺbkou a vonkajšia vzdialenosť receptívneho poľa je v tejto oblasti obyčajne menšia ako 20 cm. Všetky pohybové smery sú reprezentované. Rôzne bunky preferujú pohyb zľava, sprava, zhora, zdola a taktiež pohyb objektov oproti a od opice. Smerová preferencia pohybu vizuálnej a dotykovej modality je rovnaká pre jednotlivé neuróny pre 80% neurónov oblasti PZ.

Neuróny oblasti VIP sú bimodálne a reagujú hlavne na dotykové podnety na tvári a vizuálne podnety blízko tváre [6,9]. Väčšina týchto buniek je, podobne ako v oblasti PZ, orientačne selektívna. Niektoré z týchto buniek odpovedajú na blízke podnety vzdialené iba pár centimetrov od tváre, iné zas aj na vzdialenejšie podnety. Podobne ako v oblasti PZ, orientačná citlivosť pre jednotlivé modality je pre jednotlivé neuróny zhodná. Niektoré neuróny tejto oblasti sú dokonca trimodálne a reagujú aj na sluchové podnety. Odlišnosťou týchto neurónov sú rôzne súradnicové systémy, na ktoré sú ukotvené jednotlivé modality neurónov. V oblasti PZ sú všetky neuróny ukotvené na súradnicový systém spojený s časťou tela, ktorá je reprezentovaná danými receptívnymi poľami. V oblasti VIP je zraková modalita ukotvená na súradnicový systém so stredom definovaným smerom pohľadu, čo znamená, že ak sa pohne oko, zmenia sa aj koordináty objektov vzhľadom na tento stred. Dotyková modalita je ukotvená na tú časť tváre, ktorú pokrýva receptívne pole daného neurónu.

Z hľadiska výskumu krosmodálnych vplyvov sluchovej a vizuálnej pozornosti je dôležitá oblasť LIP. Neuróny tejto oblasti bimodálne reagujú na vizuálne a sluchové podnety, pričom táto oblasť je zodpovedná za plánovanie pohybu očí. Oblasť LIP parietálnej kôry reprezentuje prepojenie medzi vizuálnym systémom a sluchovým systémom. Problémom pri spracovaní bimodálnych vstupov sú v tomto prípade rôzne súradnicové systémy, ktoré reprezentujú priestor. Sluchová priestorová informácia je ukotvená v súradnicovej sústave definovanej stredom hlavy a polohou uší, zatiaľ čo vizuálna informácia je ukotvená v súradnicovej sústave definovanej stredom oka a

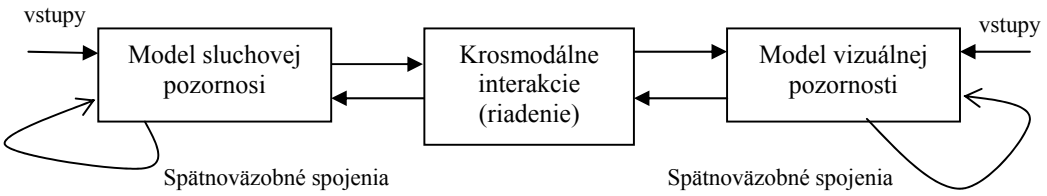
smerom pohľadu. Tieto dve reprezentácie sa musia zjednotiť, ak má byť výsledná reprezentácia priestoru konzistentná. Výsledky predošlého výskumu nasvedčujú, že priestorové koordináty sluchového podnetu sú prekódované do súradnicovej sústavy ukotvenej na stred oka [4].

5 Existujúce modely pozornostného riadenia a možnosti ich rozšírenia

Modely krosmodálnych vplyvov riadenia pozornosti musia zahŕňať výpočtové mechanizmy reprezentácie priestorovej informácie o podnete v jednotlivých modalitách spolu s pozornostnými a krosmodálnymi mechanizmami. Tieto mechanizmy sú najdetailnejšie popísané pre vizuálne vnímanie. Výsledky experimentálnych a teoretických prác naznačujú, že vizuálny kortex venuje podstatné procesné zdroje generovaniu 3-rozmerných reprezentácií perceptuálnych hraníc a povrchov, a špecificky generovaniu reprezentácií, ktoré dokážu separovať figúry od ich pozadia a doplniť reprezentácie čiastočne zakrytých objektov. Grossberg [13] navrhol, že tieto reprezentácie hraníc a povrchov sú formované v prúdoch, ktoré spracúvajú informácie vo vizuálnych kortikálnych oblastiach V1 až V4. Tieto reprezentácie potom postupujú do vyšších oblastí mozgu, najmä do inferotemporálneho kortexu (IT) vo ventrálnej dráhe, kde sú kategorizované alebo zjednotené do objektových reprezentácií. Naviac, všetky kortikálne oblasti sú navzájom prepojené spätnými väzbami a väzbami na ďalšie oblasti. Pre popis unimodálnej vizuálnej reprezentácie v rámci modelu krosmodálneho spracovania informácií je najdôležitejší popis dopredných a spätných interakcií medzi oblasťami V1 a V4. Vhodným kandidátom pre popis pozornostného vizuálneho riadenia je Teória adaptívnej rezonancie [11,16], a špecificky model vizuálnych štruktúr založený na tejto teórii nazvaný LaminART [13]. Tento model navrhuje neurálne mechanizmy umožňujúce vyvíjajúcim sa kortikálnym obvodom stabilizovať sa, a to využitím vlastností ich samoorganizovaných obvodových interakcií spolu s implementáciou komplexnej dynamiky neurónov.

Alternatívnym jednoduchším modelom vizuálnej pozornosti, ktorý je možné využiť pre skúmanie krosmodálnych pozornostných vplyvov vo vizuálnom kortexe, je model navrhnutý Andogom [1]. Tento model popisuje interakcie medzi vrstvami V1 – V4 a inferotemporálnou oblasťou (IT) pre získanie priestorovej informácie o objekte pozornosti. Model vychádza z hypotézy o objektovo orientovanej pozornosti, pričom je schopný simulovať jednak selekciu v zmysle objektovo založenej teórie a jednak proces sledovania objektov.

Na popis aktivity sluchovej dráhy môže byť použitý model periférneho sluchového systému [25] skombinovaný s modelom binaurálneho sluchového spracovania v mozgovom kmeni od Colbulrna (1977).



Obrázok 2. Schematické zobrazenie štruktúry prepojenia modelov

Pri popise krosmodálnych interakcií je potrebné poznať mechanizmy krosmodálneho riadenia, ktoré v súčasnosti nie sú detailne preskúmané. Riadenie exogénnej pozornosti je možné popísať ako pár uzavretých spätnoväzobných slučiek, ktoré navzájom interagujú prostredníctvom riadiaceho bloku (viď Obr. 2).

6 Záver

Z uvedeného prehľadu vyplýva, že napriek jednoznačnému významu strategického riadenia priestorovej pozornosti pre každodenné správanie sa človeka, sú dnes jej mechanizmy pochopené len čiastočne a to ako na neurofyziologickej, tak aj na behaviorálnej úrovni. Taktiež existuje len niekoľko málo výpočtových modelov, ktoré sa snažia tieto mechanizmy teoreticky popísať. Keďže podobné mechanizmy môžu byť užitočné aj pre umelé systémy, ktoré by mali byť schopné orientovať sa v priestore na základe integrácie vizuálnych a sluchových podnetov, je porozumenie riadeniu strategickej pozornosti jednou z dôležitých výziev pre kognitívnu neurovedu.

Poďakovanie: Tento príspevok vznikol za podpory grantovej agentúry VEGA v rámci grantovej úlohy 1/3134/06 - Unimodálne a krosmodálne riadenie strategickej pozornosti: behaviorálne experimenty a modelovanie.

Literatúra

[1] Andoga, R.: *Modelovanie objektovo založenej pozornosti použitím umelých neurónových sietí*. Diplomová práca. FEI TU Košice 2003.
[2] Broadbent, D. E.: *Perception and Communication*. Pergamon, Elmsford, New Jersey, 1958.
[3] Cave, K. R., Pashler, H.: Visual selection mediated by location: electing successive visual objects. *Perception & Psychophysics* **57** (1995) 421-432.
[4] Cohen, Y. E., Andersen, R. A.: A common reference frame for movement plans in the posterior parietal cortex. *Nature Reviews Neuroscience* **3** (2002) 553–562.
[5] Colby, C. L.: *Spatial Perception*. In: R. A. Wilson and F. Keil (eds.) MIT Encyclopedia of Cog. Sciences. MIT Press 1999.
[6] Colby, C. L., Duhamel J. R., Goldberg M. E.: Ventral intraparietal area of the macaque: anatomical location and visual response properties. *J Neurophysiol*. **69** (1993) 902–914.
[7] Critchley, M.: *The parietal lobes*. Hafner 1966.
[8] Driver, J., Spence, Ch.: *Spatial synergies between auditory and visual attention*. In Attention and performance: conscious and nonconscious information processing (Ed. C. Umilta and M. Moscovitch) MIT Press Vol. **15** (1994), p. 311-31.
[9] Duhamel, J. R., Colby, C. L., Goldberg, M. E.: Ventral intraparietal area of the macaque: congruent visual and somatic response properties. *J Neurophysiol.* **79** (1998) 126–136.
[10] Duncan, J.: Selective attention and the organization of visual information. *Journal of Experimental Psychology: General* **113** (1984) 501-517.

[11] Grossberg, S.: Adaptive pattern classification and universal recoding: I. parallel development and coding of neural feature detectors. *Biological Cybernetics* **23** (1976) 121-134.
[12] Grossberg, S.: How does a brain build a cognitive code? *Psychological Review* **87** (1980) 1-51.
[13] Grossberg, S.: 3D vision and gure-ground separation by visual cortex. *Perception and Psychophysics* **55** (1994) 48-120.
[14] Cherry, E. C.: Some experiments upon the recognition of speech with one and two ears. *Journal of the Acoustical Society of America* **25** (1953) 975-979.
[15] Kanwisher, N., Driver, J.: Objects, attributes, and visual attention: Which, what, and where. *Current directions in Psychological Science* **1** (1992) 26-31.
[16] Kopčo, N., Carpenter, G. A.: *Graded Signal Functions for ARTMAP Neural Networks*, In Sincak et al. (Eds.) The State of the Art in Computational Intelligence (Collection of papers presented at the European Symposium on Computational Intelligence, Kosice, Slovakia, Aug 30 - Sept 1, 2000). Physica-Verlag. p. 9-14.
[17] Kopčo, N., Ler, A., Shinn-Cunningham, B. G.: Effect of auditory cuing on azimuthal localization accuracy. *J Acoust Soc America* **109** (2001) 23-77.
[18] LaBerge, D.: *Attentional processing*. Cambridge, MA: Harvard University Press 1995.
[19] Mountcastle, V. B., Lynch, J. C. G. A., Georgopoulos, A., Sakata, H., Acuna, C.: Posterior parietal association cortex of the monkey: Command functions for operations within extrapersonal space. *Journal of Neurophysiology* **38** (1975) 871-908.
[20] Shimojo, S., Miyauchi, S., Hikosaka, O.: Visual attention field can be assessed by illusory line motion sensation. *IOVS* **33** (1992) 13-54.
[21] Snyder, L.H., Batista, A.P., Andersen, R.A.: Coding of intention in the posterior parietal cortex. *Nature* **386** (1997) 167–170.
[22] Soto-Faraco, S., Spence, C.: Modality-specific auditory and visual temporal processing deficits. *Quarterly Journal of Experimental Psychology A* **55** (2002) 23-40.
[23] Spence, Ch., Driver, J.: *Crossmodal Space and Crossmodal Attention*. Oxford University Press 2004.
[24] Treisman, A.M., Davies, A.: *Divided attention to ear and eye*. In Attention and Performance (ed. S. Kornblum), Vol. 4, p. 101-117. New York: Academic Press 1973.
[25] Zhang, X., Heinz, M.G., Bruce, I.C., Carney, L.H.: A phenomenological model for the responses of auditory-nerve fibers. I. Nonlinear tuning with compression and suppression. *J. Acoust. Soc. Am.* **109** (2001) 648-670.

Kurt Gödel: život, výsledky a jejich význam

Libor Běhounek[1]

Abstrakt. Kurt Gödel bývá považován za nejvýznamnějšího logika 20. století, zejména díky svým větám o neúplnosti. Vedle základních biografických a bibliografických údajů článek podává přehled Gödelových nejdůležitějších výsledků, jimiž zasáhl do matematické logiky, teorie množin, filozofie matematiky a fyziky. Zvláštní pozornost je věnována výkladu vět o neúplnosti, jejich demytizaci a důsledkům pro informatiku, umělou inteligenci a filozofii mysli.

1 Život Kurta Gödela

Kurt Gödel se narodil 28. dubna 1906 v Brně jako syn spolumajitele prosperující textilní továrny. V Brně žil až do roku 1923, kdy úspěšně maturoval na zdejším německém gymnáziu.[2] V následujícím roce se přestěhoval do Vídně a nastoupil ke studiu na Vídeňské univerzitě. Zpočátku nebyl rozhodnut mezi matematikou a teoretickou fyzikou, postupně však převážil jeho zájem o matematiku a logiku. Od roku 1926 byl jako Hahnův velmi nadaný student zván na pravidelné schůzky prestižního Vídeňského kruhu – vedoucí skupiny logického pozitivizmu. Přestože filozoficky byl v té době již vyhraněným platonistou, do debat Kruhu nevstupoval a své filozofické přesvědčení, ostře kontrastující s názory Kruhu, si nechával pro sebe.

V roce 1929 studium dokončil odevzdáním dizertace, již tvořil důkaz úplnosti predikátové logiky prvního řádu (publikována byla v [4]). Tento výsledek prezentoval v krátkém referátu na významné konferenci o základech matematiky v Královci v říjnu 1930.[3] Protože byl výsledek o úplnosti v tehdejším vědeckém kontextu očekávaný, nevzbudil referát mezi přítomnými větší rozruch. V plenární diskuzi poslední den konference Gödel krátce zmínil, že lze udat příklady pravdivých nedokazatelných vět aritmetiky. Jediný z účastníků, kdo toto oznámení neignoroval, byl John von Neumann: v kuloárech si nechal od Gödela vysvětlit základní myšlenku věty a jejího důkazu, který měl Kurt Gödel zřejmě již několik měsíců hotov, a v Princetonu o ní zpravil své kolegy. Gödel první větu o neúplnosti publikoval následujícího roku 1931 [5] a podal ji jako svou habilitační práci, na jejímž základě se v roce 1933 stal na Vídeňské univerzitě soukromým docentem.

V akademickém roce 1933–34 byl pozván do Princetonu, kde o svých výsledcích přednášel na nově otevřeném Ústavu pro pokročilá studia; zde se také seznámil s Albertem Einsteinem, který sem krátce předtím utekl před nacizmem. Po návratu do Vídně se roku 1938 oženil a vzápětí odjel (bez manželky) znovu do Ameriky, kde přednášel o svých výsledcích v teorii množin.

---

[1] Ústav informatiky AV ČR, Pod Vodárenskou věží 2, 182 07 Praha 8 a Filozofická fakulta Univerzity Karlovy v Praze, Katedra logiky, Celetná 20, 116 42 Praha 1. E-mail: behounek@cs.cas.cz.

[2] K českému prostředí zřejmě příliš vřelý vztah neměl (zejména po vzniku Československa, jehož občanem byl v letech 1918–1929): na rozdíl od spolužáků nepoužíval ani zlomky češtiny, a když byla po druhé světové válce vila jeho matky podle československo–rakouské dohody vyvlastněna za desetinu její odhadní ceny, vnímal to jako osobní křivdu ([3], [14]).

[3] Přítomni byli Hilbert, Carnap, Heyting, von Neumann a další.

V létě 1939 se vrátil do Vídně, aby zde obhajoval svůj nárok na docenturu, podléhající nyní novým pravidlům nacistického Německa. Na podzim se chtěl vrátit do Princetonu, vyřízení jeho žádosti o docenturu se však pro jeho naprostou apolitičnost a předchozí styky s židovskými kolegy protahovalo. Po začátku války se Gödel začal obávat odvedení do armády (jednou byl navíc napaden skupinou nacistických výtržníků, kteří si jej spletli se Židem); koncem roku se proto raději rozhodl pro emigraci. Po intervenci amerických kolegů mu bylo na přelomu roku povoleno vycestovat, a to přes Rusko a Japonsko: do Princetonu tak s manželkou dorazil až na jaře 1940. V Princetonu se potom zdržoval až do konce života, nejprve jako dočasný člen Ústavu pro pokročilá studia (s prodlužovanou roční smlouvou), od roku 1946 pak jako jeho stálý člen a od roku 1953 konečně jako profesor. Evropu již nikdy nenavštívil; roku 1948 získal americké občanství a v průběhu dalších let čestné doktoráty prestižních amerických univerzit (Harvardovy, Yalovy ad.) a různá jiná ocenění (Einsteinovu cenu, členství v národních akademiích a společnostech aj.).

Gödel žil vždy velmi uzavřeným životem. Společnosti se převážně vyhýbal a byl spíše plachý, v některých směrech však dosti umanutý. Od dětství trpěl neurotickými obtížemi, hypochondrií (po přestálé revmatické horečce ve věku šesti let se po celý život mylně domníval, že má poškozené srdce) a dalšími psychickými poruchami, jež různé okolnosti jeho života postupně prohlubovaly. Mezi prvními dvěma pobyty v Americe prodělal dvojí nervové zhroucení a musel nějaký čas pobývat v sanatoriu. S věkem se jeho uzavřenost zvětšovala a v Princetonu se mimopracovně stýkal jen s několika přáteli – zejména s Einsteinem, po jeho smrti roku 1955 pak hlavně s logiky Hao Wangem a G. Kreiselem, vídeňským ekonomem Morgensternem a nemnoha dalšími; děti neměl. Po roce 1958 přestal své výsledky publikovat, přestože stále pracoval; nevedl ani žádné studenty. Paranoidní strach z otravy jídlem vedl k jeho dlouhodobé podvýživě; během několikaměsíčního pobytu jeho ženy v nemocnici na podzim 1977 pak vyhladověl k naprosté vyčerpanosti, a ani pobyt v nemocnici, kam jej nechala po svém návratu odvézt, jej již nezachránil; zemřel 14. ledna 1978.[4]

2 Gödelovy výsledky a přínos

V této a následující části podáme základní přehled Gödelových výsledků a jejich stručné zhodnocení. Vzhledem k obsáhlosti tématu a stavu jeho prozkoumání může jít pochopitelně pouze o shrnující a přehledový popis; podrobnosti a další informace lze najít v odkazované literatuře. Její výběr se částečně řídil snadnou dostupností pro české a slovenské čtenáře: z velkého množství relevantní literatury proto byly přednostně použity práce vydané česky nebo slovensky a zdroje přístupné na internetu.

2.1 Logika

Nejznámějším Gödelovým výsledkem v logice jsou bezesporu jeho věty o neúplnosti, vnímané často jako ničím nepřekonatelný výkon. I podle Gödelova osobního názoru (citovaného v [14]) však bylo v letech 1930–31 otázkou měsíců, kdy by na tyto věty

[4] Podrobnější životopisné údaje lze najít např. v [14] či [3].

narazil někdo jiný.[5] Jeho výsledky proto není třeba titanizovat;[6] přišel však s nimi jako první a ve velmi propracované a čisté formě. Vedle vět o neúplnosti aritmetiky podal i první důkaz úplnosti prvořádové logiky a významně přispěl k teorii i metateorii intuicionistické logiky a matematiky.

2.1.1 Věta o úplnosti prvořádové logiky

Věta o úplnosti nějakého formálního systému říká, že všechny jeho *sémanticky platné* formule v něm lze *syntakticky odvodit.* Pro formální systém logiky prvního řádu jde tedy o větu, že všechny logicky platné formule jazyka prvního řádu jsou dokazatelné (v dané axiomatice této logiky). Dnešní axiomatika prvořádové logiky byla navržena v roce 1928 v Hilbertově a Ackermannově knize *Grundzügen der theoretischen Logik.* V následujícím roce Gödel kladně vyřešil (v ní formulovaný) problém její úplnosti, neboli dokázal následující větu:

Věta (o úplnosti): Každá tautologie klasické prvořádové logiky je dokazatelná v Hilbertově–Ackermannově systému axiomů.

Spolu se (snadnou) větou o korektnosti této axiomatiky (jež říká, že dokazatelné jsou v ní *pouze* tautologie prvořádové logiky) Gödelův výsledek znamená, že pojem logické pravdivosti v jazyce prvního řádu přesně odpovídá jednoduchému pojmu formální („mechanické“) dokazatelnosti v určitém formálním axiomatickém systému.

Význam tohoto výsledku je značný: ukazuje, že pro prvořádová tvrzení lze pojem logické pravdivosti – tedy pravdivosti ve všech (čili nekonečně mnoha) strukturách (i nekonečných) – redukovat na dokazatelnost pomocí konečné mechanické procedury ve finitně zadaném axiomatickém systému. Tato věta se zdála podporovat formalistickou doktrínu Hilbertovy školy, požadující redukovat veškerou matematiku, zabývající se i nekonečnými strukturami, na jednoduché manipulace s konečně zadanými objekty. (Její neproveditelnost ukázaly až věty o neúplnosti.)

Z hlediska obecného dosahu této věty je třeba mít na paměti, že se týká pouze logiky prvního řádu a platí vlastně právě díky omezenosti jejích výrazových prostředků. Jazyk prvořádové logiky obsahuje vedle výrokových spojek (konjunkce, disjunkce, negace, implikace, ekvivalence) pouze proměnné pro individua nějakého daného univerza, jména pro predikáty nějakého daného jazyka[7] a možnost kvantifikace přes individuové proměnné. Na rozdíl od logiky druhého a vyšších řádů neobsahuje *proměnné* (nýbrž jen jména) pro predikáty a funktory, a neumožňuje tedy přes ně ani kvantifikovat.

Unární predikáty označují vlastnosti individuí, reprezentované podmnožinami univerza; binární a více-ární predikáty vyjadřují vztahy mezi těmito individuy a jsou reprezentovány relacemi příslušné arity na univerzu. V logice prvního řádu tedy můžeme mluvit o *jednotlivých* predikátech, pro něž máme v daném jazyce jména – např. v jazyce aritmetiky o predikátu „být menší“ <. Nemůžeme však přes predikáty (neboli podmnožiny a relace) kvantifikovat. V prvořádovém jazyce tedy nelze vyjádřit ani tak důležité věty

[5] Šlo tehdy o aktuální témata a lhářovské motivy byly již po čtvrtstoletí předmětem intenzivního zkoumání; Zermelo si patrně byl vět o neúplnosti vágně vědom a výzkumy dalších logiků (Post, Tarski, Skolem) k nim měly tematicky velmi blízko.

[6] Jak činí např. [3]; střízlivější pohled nabízejí [14] a [13].

[7] Případně i pro funktory; protože však mohou být definovány pomocí predikátů, nechme je, stejně jako Gödel, pro jednoduchost stranou.

jako *Každá množina přirozených čísel má nejmenší prvek* či *Každá omezená množina reálných čísel má supremum.* Tyto věty obsahují kvantifikaci přes podmnožiny, jsou tedy druhořádové.

Úplnost prvořádové logiky tedy neznamená redukci matematické pravdivosti na dokazatelnost, neboť mnoho důležitých matematických vět je druhořádových. Gödelovy věty o neúplnosti ukazují, že tato redukce obecně skutečně není možná.

2.1.2 Věty o neúplnosti

Tzv. první Gödelova věta o neúplnosti konstatuje neúplnost určitého druhu formálních teorií. V jedné z nejstručnějších verzí může být formulována takto:

Věta (první o neúplnosti): Každá bezesporná rekurzívně axiomatizovaná teorie obsahující Peanovu aritmetiku je neúplná.

Peanova aritmetika PA je prvořádovou teorií obsahující tři axiomy pro funkci následníka neboli přičtení jedničky (nula není následníkem žádného čísla, každé nenulové číslo je následníkem nějakého čísla, a čísla si jsou rovna, jsou-li si rovni jejich následníci); rekurzívní definice sčítání pomocí následníka ($x+0 = x$, $x+\text{násl}(y) = \text{násl}(x+y)$) a násobení pomocí sčítání ($x\cdot 0 = 0$, $x\cdot\text{násl}(y) = x\cdot y+x$); a axiom indukce pro každou prvořádovou aritmetickou vlastnost φ (tedy tvrzení, že má-li vlastnost φ nula a s každým číslem i jeho následník, pak ji už mají všechna čísla).[8]

Vlastnost *rekurzívnosti* lze považovat za ekvivalentní algoritmizovatelnosti pomocí Turingova stroje (v praxi tedy počítače). Teorie je rekurzívně axiomatizovaná, má-li rekurzívní sadu axiomů a odvozovacích pravidel. Pouze takové teorie lze rozumně chápat jako „finitně“ zadané, neboť pouze u nich máme konečný algoritmus, jak poznat jejich axiomy.

Neúplnost teorie znamená, že existuje tvrzení, které v ní nelze ani dokázat ani vyvrátit. Je-li daná teorie navíc korektní (tj. jsou-li v ní dokazatelné pouze pravdivé věty aritmetiky), pak existují *pravdivá* aritmetická tvrzení, která v ní nejsou dokazatelná. V obou případech Gödelův důkaz umožňuje tato tvrzení přímo sestrojit, tj. ukázat konkrétní takovou formuli (nazývanou *Gödelova sentence*).

Myšlenka důkazu spočívá v zakódování syntaxe formální teorie pomocí čísel. Každé formuli je tak algoritmicky přiřazen její číselný kód, tzv. Gödelovo číslo. Tím jsou syntaktické vlastnosti vyjádřeny v jazyce aritmetiky a lze o nich *v aritmetice* dokazovat potřebná tvrzení. Předpoklad úplnosti teorie lze pak obratem podobným Cantorově diagonální metodě přivést ke sporu. Přes jasnou základní myšlenku je provedení všech detailů důkazu poměrně pracné a vyžaduje značné množství definovaných pojmů a předběžných lemmat.[9]

---

[8] Modifikací důkazu lze získat několik variant Gödelových vět, lišících se silou předpokladů a závěru; první věta o neúplnosti například platí již pro tzv. Robinsonovu aritmetiku, tj. výše uvedenou axiomatiku *bez* axiomů indukce. Rozbor těchto variant však není předmětem tohoto článku; detaily lze nalézt v [21], [23] či [18]. Gödelova původní formulace používala místo bezespornosti o něco silnější předpoklad tzv. ω-bezespornosti, na pouhou bezespornost zesílil větu J. B. Rosser.

[9] Omezený rozsah tohoto článku nedovoluje uvést další podrobnosti; pro detaily viz česky [21] či [23], na webu např. [18] či přímo [5]; přehled důkazu lze najít v [13] či [25], jeho populární výklad např. v [3], [16], [11], [19], [20].

V důsledku Gödelovy věty je značná část zajímavých dostatečně silných matematických teorií principiálně neúplná. Věta navíc tvrdí, že teorie zůstane neúplnou, ať ji rozšíříme jakýmkoli rozumným způsobem – tj. tak, aby zůstala bezesporná (spor by teorii zcela znehodnocoval, neboť lze přes něj odvodit cokoliv) a rekurzívně axiomatizovaná (abychom byli schopni algoritmicky poznat její axiomy).

Důsledkem vhodného tvaru první Gödelovy věty je tzv. druhá věta o neúplnosti, formulovatelná např. takto:[10]

Věta (druhá o neúplnosti): Žádná bezesporná rekurzívně axiomatizovaná teorie T obsahující Peanovu aritmetiku nedokazuje formuli Con(T) vyjadřující její formální bezespornost.

Rigorózní výklad formule Con(T) a přesná formulace její souvislosti s bezesporností teorie T přesahuje rámec tohoto článku.[11] Lze ji však v určitém smyslu chápat jako formální protějšek bezespornosti teorie T a větu interpretovat tak, že v dostatečně silné bezesporné rekurzívní teorii nelze dokázat její vlastní bezespornost. Výkladu důsledků a významu obou vět o neúplnosti věnujeme samostatný oddíl 3.

2.1.3 Nekonečněhodnotovost intuicionistické logiky

Na počátku dvacátého století matematik L. E. J. Brouwer vystoupil s požadavkem konstruktivního vedení důkazů v matematice. Jeho přístup, nazvaný intuicionizmus, získal v první polovině dvacátého století (jakožto jedna z reakcí na krizi v základech matematiky způsobenou paradoxy teorie množin a infinitní matematiky) poměrně velký ohlas. Intuicionistická matematika nepřipouští důkazy sporem, neuznává zákon dvojité negace $A \leftrightarrow \neg\neg A$ ani zákon vyloučení třetího $A \vee \neg A$ apod. Logiku používanou v intuicionistické matematice axiomatizoval koncem dvacátých let A. Heyting. Klasická výroková logika je dvouhodnotová – lze ji zavést pomocí pravdivostních tabulek se dvěma hodnotami 0 a 1. Několik let nebylo jasné, zda nelze podobně chápat intuicionistickou logiku jako troj- či vícehodnotovou. Bylo nalezeno několik sad vícehodnotových pravdivostních tabulek, vůči nimž je intuicionistická logika korektní (tj. při počítání podle těchto tabulek vycházejí intuicionisticky platné formule jako pravdivé). Vždy se však ukázalo, že není vůči nim úplná (tabulky „přegenerovávaly“ – byly podle nich pravdivé i některé intuicionisticky neplatné formule). Ve svém dvoustránkovém článku [6] Gödel demonstroval marnost těchto snah – pomocí vtipného obratu dokázal, že intuicionistická logika nemůže být zadána pomocí žádné pravdivostní tabulky s konečně mnoha hodnotami.

Pro účely důkazu tohoto tvrzení Gödel přidal k intuicionistické logice axiom $(A \rightarrow B) \vee (B \rightarrow A)$. Přestože Gödel výslednou logiku ani nepojmenoval a nijak ji dále nepoužil, byla to první zmínka o důležitém logickém systému, který později důkladně prozkoumal M. Dummett. Po vzniku fuzzy logiky se ukázalo, že toto rozšíření intuicionistické logiky je zároveň jednou z nejdůležitějších fuzzy logik, v níž je konjunkce realizována jako minimum pravdivostních hodnot obou argumentů v intervalu [0, 1]. Vedle důkazu nekonečněhodnotovosti intuicionistické logiky se tak Gödel stal maně objevitelem jedné z nejdůležitějších fuzzy logik, dnes zvané Gödelova či Gödelova–Dummettova.

---

[10] Gödelova původní formulace se opět v detailech liší.

[11] Netriviálnost tohoto vztahu je vidět i z toho, že dle druhé věty o neúplnosti je teorie PA + ¬Con(PA), tedy Peanova aritmetika s předpokladem své vlastní formální *spornosti,* bezesporná.

2.2 Teorie množin

Teorie množin je spolu s logikou součástí „základů matematiky“ a zejména v první polovině dvacátého století ji bylo možno považovat za obor „sousedící“ s logikou. Je proto přirozené, že Gödel byl aktivní i v něm.

2.2.1 Nevyvratitelnost hypotézy kontinua a axiomu výběru

V roce 1938 Gödel jedním tahem z poloviny vyřešil dva dlouho otevřené významné problémy v teorii množin:

Koncem 19. století zakladatel teorie množin Georg Cantor vyslovil hypotézu, že každá podmnožina reálných čísel je buďto spočetná (tj. existuje její očíslování přirozenými čísly), nebo má mohutnost kontinua (tj. existuje její vzájemně jednoznačné zobrazení na množinu R všech reálných čísel); jiné podmnožiny R totiž nemohl nalézt. Přes několik částečných výsledků potvrzujících, že podmnožiny R nejrůznějších vlastností jeho hypotézu splňují, obecný důkaz Cantor ani nikdo jiný nebyl schopen podat. Tato *hypotéza kontinua,* v kardinální aritmetice zapsatelná jako $2^{\aleph_0} = \aleph_1$, byla zobecněna na ještě silnější hypotézu stanovující $2^{\aleph_\alpha} = \aleph_{\alpha+1}$ pro každé ordinální číslo α – tedy tvrzení, že mezi mohutností *žádné* nekonečné množiny a její potence (množiny všech jejích podmnožin) již nejsou žádné další mohutnosti. Ani tuto silnější hypotézu (GCH) se nikomu nepodařilo vyvrátit ani dokázat.

V souvislosti s axiomatizací teorie množin (v reakci na nalezené paradoxy) na počátku dvacátého století se ukázalo, že mnoho důkazů v matematice používá obrat, kdy se současně vybere nekonečně mnoho nespecifikovaných prvků. Předpoklad, že to lze, precizovaný a formulovaný jako *axiom výběru* (AC), zní v jedné ze svých mnoha ekvivalentních podob takto: ke každé množině *A* neprázdných množin existuje množina, která má s každou množinou z *A* právě jednoprvkový průnik. Bez axiomu výběru nebylo možno dokázat mnoho důležitých výsledků, mimo jiné i tak „očividná“ tvrzení, jako že kartézský součin systému neprázdných množin je neprázdný; jeho nekonstruktivní povaha – nedává totiž žádný předpis, jak výběrovou množinu *sestrojit* – jej však činila pochybným nejen v očích intuicionistů. Velkou snahou matematiků po více než tři desítky let proto bylo buď tento axiom dokázat z ostatních axiomů teorie množin (čímž by byl ospravedlněn), nebo ukázat jeho nezávislost na ostatních axiomech teorie množin (čímž by byl alespoň vyjasněn jeho status, podobně jako u pátého Eukleidova postulátu).

Gödelův výsledek oznámený v [7] používá konstrukci popisující tzv. *konstruovatelné množiny.*[12] Gödel ukázal, že univerzum všech konstruovatelných množin (označované L) splňuje všechny axiomy teorie množin, a navíc v něm platí jak axiom výběru, tak i zobecněná Cantorova hypotéza kontinua. Ukázal tedy, že AC i GCH jsou s obvyklými Zermelovými–Fraenkelovými axiomy teorie množin konzistentní, nedají se tedy pomocí nich vyvrátit. Druhou polovinu důkazu jejich nezávislosti, tj. jejich nedokazatelnost, ukázal v šedesátých letech P. Cohen.

[12] Tyto množiny lze charakterizovat jako počitatelné nekonečnými Turingovými stroji v nekonečném čase (kde páska i takty mohou mít libovolný transfinitní ordinální typ). Gödelova definice však byla jiná a opírala se o ordinálně iterovanou definovatelnost pomocí formulí.

Konstrukce univerza L hrála pak důležitou roli i v dalším vývoji teorie množin. Toto univerzum je prototypem jejích tzv. *vnitřních modelů* (vnitřních proto, že L je poduniverzem univerza *všech* množin, označovaného V; navíc obsahuje všechna ordinální čísla). Pomocí konstrukcí vnitřních modelů lze ukázat řadu dalších výsledků o nedokazatelnosti či nevyvratitelnosti některých principů teorie množin. Velmi podrobný popis Gödelovy konstrukce včetně všech důkazů a aplikací těchto metod lze najít v [22]; dobrý celkový obrázek lze získat i z přehledu v [1], [14] či [13].

Přijetí předpokladu, že každá množina je konstruovatelná (tedy axiomu V=L), řeší nejenom problém platnosti AC a GCH, ale i mnoho dalších problémů teorie množin. Gödel sám ihned ukázal, že z něj plynou odpovědi na některé tehdy otevřené problémy deskriptivní teorie množin (zda existují množiny reálných čísel s jistými vlastnostmi). Protože je však univerzum L na množiny velice „chudé“ (v jistém smyslu nejchudší možné), není axiom V=L v teorii množin běžně přijímán a L je obvykle chápáno jen jako velmi důležitý vnitřní model. Takto L chápal i sám Gödel; přes svůj výsledek o její nevyvratitelnosti byl totiž přesvědčen, že Cantorova hypotéza ve skutečnosti neplatí.[13]

2.2.2 Gödelova–Bernaysova–von Neumannova axiomatizace teorie tříd

Vedle tohoto významného příspěvku žije Gödelovo jméno v každodenním hovoru množinových teoretiků i v názvu jedné z nejvýznamnějších axiomatik teorie množin, označované jako Gödelova–Bernaysova, GB (též NBG, von Neumann–Bernays–Gödel). Na rozdíl od staršího Zermelova–Fraenkelova systému ZF jsou jeho objekty nikoli množiny, nýbrž tzv. *třídy*. Množiny jsou definovány jako ty třídy, které jsou prvky jiných tříd. Množinami v zásadě mohou být jen „malé“ třídy; tím jsou ošetřeny paradoxy teorie množin propukající jinak u „nadměrně velkých“ tříd, jako je třída všech množin V. Díky opatrné formulaci schématu axiomů definování tříd, požadující kvantifikaci pouze přes *množiny* (nikoli třídy), je tato teorie stejně silná jako ZF a lze ji chápat jako její notační variantu (třídy lze zavést v ZF jako zkratky za formule, viz [22] či [1]); jejich rozdíly však hrají roli v metamatematických zkoumáních: GB je např., narozdíl od ZF, konečně axiomatizovatelná. Podrobný popis teorie GB a jejích vlastností je uveden v [22] (pro stručný přehled viz např. [25], heslo „NBG“).

2.3 Fyzika a filozofie

Dalším oborem, do nějž Gödel – sice jen jednou, zato však významně – zasáhl, byla obecná teorie relativity. Celý život se rovněž zabýval problémy filozofie matematiky; jeho eseje o ní však byly publikovány až z jeho pozůstalosti. Vedle toho se široké publicitě teší jeho spíše kuriózní verze důkazu boží existence. Těmito tématy proto přehled jeho hlavních výsledků uzavřeme.

2.3.1 Filozofie matematiky: Gödelův program

Gödel byl od svých studií přesvědčeným platónským realistou. Ve filozofii matematiky tento názor znamená přesvědčení, že matematické objekty existují nezávisle na našich

[13] Tento jeho názor souvisí s jeho platonistickým přesvědčením ve filozofii matematiky, o němž pojednává sekce 2.3.1.

axiomech a metodách: ty je nevytvářejí, jen popisují. I když tedy podle Gödelových vět existují tvrzení, která daným formálním aparátem nelze rozhodnout, z objektivní existence matematické reality pro platonistu vyplývá, že tato nezávislá tvrzení jsou přesto buď pravdivá, nebo nepravdivá. (V tom se liší např. od intuicionisty, podle nějž tvrzení nemají pravdivostní hodnotu, dokud je nedokážeme nebo nevyvrátíme.) U složitějších nezávislých tvrzení však nemá platonista žádné vodítko, zda jsou (přes svoji nedokazatelnost a nevyvratitelnost) pravdivá či nikoli – kromě matematické intuice.

Gödel (svoji) matematickou intuici jako takové vodítko bral velmi vážně.[14] Kupříkladu Cantorovu hypotézu, jejíž nevyvratitelnost sám dokázal, považoval za nepravdivou. Jako obecný princip, který by podle něj měl rozhodnout „skutečnou pravdivost“ všech podobných nezávislých tvrzení, navrhoval postulování stále větších a větších kardinálních čísel.

Jak ukázal již Cantor, existuje nekonečné množství různých mohutností nekonečných množin. Tyto mohutnosti, označované kardinálními čísly, jsou (za předpokladu axiomu výběru) lineárně uspořádány a k větším kardinalitám lze přejít operacemi potence či sjednocení. Tzv. *nedosažitelné kardinály* jsou (volně řečeno) ta kardinální čísla, k nimž se od menších nelze pomocí potence a sjednocení dostat. Jejich existence není z axiomů teorie množin dokazatelná[15] a lze ji jedině postulovat. Postulátem tohoto druhu je již sám axiom nekonečna, zaručující existenci nekonečné spočetné množiny: ke spočetnému nekonečnu se rovněž operacemi sjednocení a potence nelze dostat od menších – konečných – množin. Konečné množiny tvoří soběstačné univerzum teorie množin bez axiomu nekonečna; ten teprve zaručuje, že toto uzavřené univerzum konečných množin „překročíme“. Podobný postulát lze přijmout i pro překročení uzavřeného univerza množin menších než nedosažitelný kardinál. Matematici tyto postuláty obvykle považují za přijatelné: rozšiřují totiž univerzum množin přirozeným způsobem směrem k větším velikostem. Předpoklad existence vhodného velkého kardinálu pak často řeší jinak nerozhodnutelné problémy teorie množin.

Gödel rovněž považoval toto rozšiřování za přirozené: „skutečné“ univerzum množin totiž podle jeho představ není nijak omezeno – ani kardinálem, jehož existenci naším (pouhým) formalizmem nelze dokázat. Byl tedy pro zesilování teorie množin jakýmkoliv prodlužováním hierarchie kardinálů.[16] Takovýmto prodlužováním kardinalit nade všechny meze měla být podle něj nakonec rozhodnuta všechna důležitá nerozhodnutelná tvrzení teorie množin.[17]

Tato strategie rozhodování nezávislých tvrzení je známa jako *Gödelův program.*[18] Přestože se s jeho pomocí skutečně dá mnoho nezávislých tvrzení rozhodnout, jsou indicie, že u některých tvrzení postulování sebevětšího kardinálu nepomůže (jedním z nich

[14] Podle některých zmínek měl pocit, že matematické univerzum vnitřním zrakem přímo nahlíží.

[15] Z existence nedosažitelného kardinálu totiž vyplývá existence modelu teorie množin, tedy její bezespornost; a ta podle druhé Gödelovy věty není v teorii množin dokazatelná.

[16] To je i důvod, proč nepovažoval za přijatelný axiom V=L: existence tzv. měřitelného kardinálu totiž vynucuje existenci nekonstruovatelných množin, tj. V≠L. Proto ani platnost GCH v L pro něj nebyla známkou její pravdivosti.

[17] K tomuto přesvědčení byl přiveden tím, že tyto předpoklady řešily nezávislá tvrzení daná jeho větami o neúplnosti – Gödelovu sentenci a formální bezespornost teorie; toto pozorování pak extrapoloval na všechna ostatní tvrzení.

[18] Viz [9], [2] či [14].

je právě Cantorova hypotéza kontinua). V současné době se tedy Gödelův program považuje pro rozhodnutí všech nezávislých tvrzení za nedostatečný.

2.3.2 Rekonstrukce ontologického důkazu boží existence

Gödel se vždy snažil pracovat na takových problémech, které mají filozofický přesah. V době kolem roku 1970 jeho kolegové věděli, že pracuje na formálním důkazu existence boží. V jeho pozůstalosti se skutečně list papíru s takovým důkazem našel. Jedná se o formalizaci Leibnizovy varianty[19] Anselmova ontologického důkazu v modální logice S5 druhého řádu.

Ve velmi stručné formulaci zní Anselmův argument takto: bůh, jakožto nejdokonalejší bytost, má z definice všechny přednosti; pro takovou bytost je existence jistě předností – musí ji tedy mít. Gödel analyzoval tento důkaz prostředky moderní logiky, aby nalezl jeho přesnou strukturu a odhalil potřebné předpoklady.

Formální důkaz nalezený v pozůstalosti je velmi stručný, nekomentovaný, a má některé nežádoucí důsledky.[20] Od jeho publikace v r. 1987 byl proto řadou autorů znovu studován, rekonstruován a opravován – ze slovenských a českých zejména P. Zlatošem [28] a P. Hájkem [10].

Na rozdíl od populárních představ, Gödelův důkaz pochopitelně není žádným nepochybným důkazem boží existence. Je pouze formální rekonstrukcí a analýzou po staletí známého ontologického důkazu, vůči němuž lze vznášet celou řadu námitek. Gödelem izolované a explicitně uvedené formální předpoklady navíc ukazují, že jejich přijetí je značně problematické, a odhalují tak de facto především slabiny „důkazu“.[21]

2.3.3 Gödelův model vesmíru

Známým Gödelovým výsledkem je rovněž jeho překvapivé řešení Einsteinových rovnic pole dávající kosmologický model, jenž umožňuje fyzickou cestou v prostoru dosáhnout libovolných časoprostorových souřadnic – tedy i cestovat v čase do minulosti. Gödel teoretické fyzice dobře rozuměl (začal ji původně studovat společně s matematikou); jeho přátelství s Einsteinem v Princetonu a jeho snaha pracovat na filozoficky relevantních matematických problémech zřejmě přispěly k hledání takto neobvyklého řešení rovnic obecné teorie relativity.

Gödelovo řešení se týká vesmíru s pečlivě zvolenou nenulovou kosmologickou konstantou, v němž jsou částice „prachu“ (což mohou být například galaxie) v určitém rotačním pohybu. Následkem toho je jeho geometrie uzavřena do sebe takovým způsobem, že obsahuje uzavřené časupodobné křivky; to vede k popsaným efektům.[22]

Článek [8] s tímto kosmologickým modelem Gödel napsal k Einsteinovým sedmdesátinám v roce 1949. Einsteina existence takových modelů jeho vlastní teorie znepokojila a soudil, že nebudou fyzikálně možné. Gödel byl však svým řešením natolik zaujat, že se zajímal o to, zda v reálném vesmíru existuje preferovaný smysl pohybu galaxií. Gödelovo řešení bylo prvním z širší třídy později nalezených řešení s uzavřenými

[19] Gödel Leibnizovy spisy velmi podrobně studoval, Leibniz byl jeho oblíbený autor.
[20] Např. tzv. kolaps modalit – jeho předpoklady způsobují, že není rozdíl mezi *A* a *nutně A*.
[21] Diskutabilní jsou zde již axiomy modální logiky S5, zejména axiom „je-li možné, že nutně *A*, pak nutně *A*“.
[22] Podrobnosti lze najít ve [25] či přímo v [8].

časupodobnými křivkami; všechna jsou však neslučitelná s pozorovanými vlastnostmi vesmíru (Gödelův vesmír například nemá Hubblovu expanzi).

3 K významu vět o neúplnosti

Stručné zhodnocení většiny Gödelových nejdůležitějších a nejznámějších výsledků jsme provedli přímo u jejich popisu v předchozí kapitole. Zde se budeme podrobněji zabývat významem jeho vět o neúplnosti, zejména z hlediska jejich filozofických interpretací. Zvláště přihlédneme k jejich případným důsledkům v informatice a umělé inteligenci.

3.1 Důsledky vět o neúplnosti v metamatematice a informatice

Principiální neúplnost dostatečně silných teorií u laiků obvykle vyvolává údiv. Z určitého podstatného hlediska však věty o neúplnosti příliš překvapivé nejsou; překvapivá je z tohoto hlediska naopak věta o *úplnosti.* Ta říká, že logickou platnost prvořádových formulí lze popsat konečnou mechanickou aplikací konečných pravidel; tedy že naše omezené *konečné* prostředky kupodivu plně popisují *celou* nekonečnou třídu *i nekonečných* matematických struktur z hlediska prvořádové platnosti. Naopak věty o neúplnosti říkají, že určitou konkrétní nekonečnou matematickou strukturu – přirozená čísla – našimi konečnými prostředky rekurzívních teorií nepopíšeme. Právě proto, že je tato struktura nekonečná, nedá se ani příliš očekávat, že to půjde – bylo by *překvapivé,* kdyby to bylo možné.[23]

Proč tedy věty o neúplnosti (a nikoli věta o úplnosti) vzbudily takový ohlas, jemuž se u laické veřejnosti těší dodnes? – Bylo to dáno zejména historickými okolnostmi. Snaha o zabezpečení matematiky pracující s nekonečnem (zejména matematické analýzy a teorie množin) proti paradoxům a mylným intuicím vedla na počátku dvacátého století k programu její formalizace – podložení „podezřelých“ tvrzení o nekonečných strukturách „solidními“ finitními metodami axiomatických systémů v rámci nedávno vzniklé formální logiky. Hlavním proponentem tohoto přístupu byl Hilbert, který sám takto formalizoval geometrii. Jedním z hlavních cílů Hilbertova programu bylo ukázat, že infinitní metody jsou redukovatelné na finitní – zejména že jejich (nejistou) bezespornost lze prokázat ryze finitními metodami. Za finitní oblast par excellence a jednu z nejjednodušších matematických disciplin se považovala (peanovská) aritmetika; jedním z hlavních cílů se tedy stalo ukázat bezespornost teorie množin (a díky ní i dalších oborů infinitní matematiky) právě pomocí aritmetiky. Gödelovy věty o neúplnosti však možnost dosažení tohoto cíle vyvrátily.

Překvapení z Gödelových vět bylo o to větší, že předchozí vývoj (kupodivu) nasvědčoval proveditelnosti Hilbertova plánu: vedle geometrie bylo krátce předtím formalizováno i několik dalších teorií (např. teorie reálných čísel) a ukázána úplnost těchto formalizací. Sama Gödelova věta o úplnosti do tohoto programu zcela zapadala (proto ani nevzbudila velkou pozornost). I v případě aritmetiky samotné již existovaly indicie pro její úplnost – byla dokázána úplnost peanovské aritmetiky bez operace násobení (tzv. Presbur-

[23] Jak trefně poznamenává Kreisel v [14], o 50 let dříve by veškerou slávu sklidila věta o úplnosti, zatímco věty o neúplnosti by pouze potvrdily původně převažující mínění, že aritmetika je na logiku neredukovatelná – už proto, že diofantické rovnice některých tvarů jsou velmi obtížné, a těžko si představit, že by šly řešit mechanicky.

gerova aritmetika) a jen se čekalo, kdy někdo vyřeší problém úplnosti aritmetiky i s násobením. V této celkem oprávněně optimistické atmosféře pak Gödel vyvrátil jak úplnou axiomatizovatelnost aritmetiky, tak možnost prokázání bezespornosti infinitní matematiky finitními prostředky (neboť podle jeho druhé věty nelze v „bezpečně bezesporné" Peanově aritmetice prokázat ani bezespornost jí samotné, natož silnější teorie množin).[24]

Hlavním důsledkem Gödelových vět tedy bylo vyvrácení původní verze Hilbertova programu, jež se díky nim ukázala přece jen jako příliš naivní. Proveditelné jsou tak jen značně oslabené verze tohoto programu (např. dokazovat lze pouze *relativní* bezespornost teorií vůči sobě navzájem apod.). Vedle opravy základního obrazu filozofie matematiky mají však Gödelovy věty v metamatematice i technický význam: využívá je např. demonstrace nedokazatelnosti existence nedosažitelných kardinálů v ZF (viz pozn. 15) či důkaz nemožnosti teorii ZF konečně axiomatizovat; podobně se jejich prostřednictvím dá ukázat prvořádová nedefinovatelnost některých struktur v určitých teoriích.[25]

Pojem rekurzívní funkce, který Gödel pro účely svého důkazu v aritmetice přesně definoval (intuitivně byl používán již dříve), formálně zachycuje její principiální algoritmickou počitatelnost. Metody použité v důkazu našly proto později značné uplatnění v teorii algoritmů a informatice. Z hlediska algoritmizovatelnosti rozpracoval gödelovské metody koncem 30. let Alan Turing (viz zvl. [24]).[26] Pomocí výpočetního modelu Turingova stroje bylo možno dát pojmu rekurzívní funkce ryze algoritmický charakter a ukázat výpočetní ekvivalenty Gödelových vět.[27] Vedle neúplnosti matematických teorií tak byla (Turingem, Churchem a dalšími) ukázána algoritmická *nerozhodnutelnost* některých úloh – neexistence algoritmu řešícího takovou úlohu obecně, pro každý vstup. V oblasti metamatematiky jde zejména o nerozhodnutelnost aritmetiky a dalších formálních teorií (např. teorie grup, okruhů ad.), včetně samotné prvořádové logiky.[28] Rozpracování gödelovských metod vedlo i např. k negativnímu řešení desátého Hilbertova problému, v němž Hilbert žádal mechanickou metodu rozpoznání řešitelnosti diofantických rovnic. Nejznámější aplikací v informatice je věta (dokázaná již v [24]) o algoritmické neřešitelnosti problému zastavení, tedy neexistence programu, který by pro každý program a jeho vstupní data rozpoznal, zda se na nich zastaví, nebo zacyklí. Podobnými obraty je možno ukázat algoritmickou neřešitelnost řady dalších úloh.

---

[24] Prokázání bezespornosti teorie v ní samé pochopitelně její bezespornost nijak neřeší, neboť bude-li sporná, dokáže svoji bezespornost také (protože dokáže cokoli). I proto Gödel sám svoji druhou větu považoval spíše za hříčku než významný důsledek. Význam má ale dokazování bezespornosti silnější teorie ve slabší, a na tento případ se jeho věta vztahuje také. Za zmínku snad stojí, že nedávno byly objeveny korektní aritmetické teorie, slabší než PA, které svou vlastní bezespornost dokazují (neumějí však např. dokázat, že součin je definován pro všechny dvojice čísel), viz [26].

[25] Např. přirozených čísel v prvořádové teorii reálných čísel. Ta je totiž úplná, a prvořádová definovatelnost přirozených čísel v ní by měla za důsledek i úplnost aritmetiky. Přirozená čísla tak lze v rámci reálných definovat pouze *druhořádovou* vlastností, např. archimédovskostí.

[26] Turing se s Gödelovými výsledky seznámil v Princetonu, kde pobýval v době mezi prvními dvěma návštěvami Gödelovými. Ačkoli se spolu nesetkali, Gödel si Turingových výsledků a jeho alternativního způsobu rozpracování svých metod velmi cenil.

[27] Gödelovu triku s aritmetickým kódováním syntaxe v nich např. odpovídá simulace Turingova stroje na univerzálním Turingově stroji – interpreteru.

[28] Ta, ač je *úplná,* je *nerozhodnutelná* – každá tautologie je v ní sice dokazatelná, ale neexistuje algoritmus, který by o každé formuli po konečném počtu kroků rozhodl, zda je tautologií či nikoli. Řečeno jazykem teorie rekurze, množina všech tautologií je sice rekurzívně spočetná, není však rekurzívní.

Věty o neúplnosti a nerozhodnutelnosti jsou ve všech těchto oblastech bezesporu významnými *výsledky;* za ještě důležitější je ale možno považovat *metody,* které do metamatematiky, teorie modelů i teorie rekurze přinesly.

3.2 Chybné interpretace Gödelových vět

Gödelovy věty bývají velmi často chybně vykládány, obvykle v důsledku neporozumění či přehlédnutí role jejich premis; jejich význam mimo metamatematiku tak bývá značně přeceňován. Naznačíme zde důvody, proč jsou zejména jejich interpretace pro přírodní vědy a pro filozofii mysli ve většině případů zcela nepřiměřené a uvedeme na pravou míru několik nejčastějších konkrétních nesprávných interpretací Gödelových vět.

Obvyklou chybou laických interpretací obou vět je jejich aplikace na situace, kterých se netýkají. Jedním z nejvulgárnějších příkladů jsou tvrzení typu: *Jak rigorózně ukázal Gödel, Bible je buďto neúplná, nebo sporná.* Protože Bible je jen sotva rekurzívní teorií obsahující Peanovu aritmetiku, věta o neúplnosti o jejích formálních vlastnostech nemluví. Jistěže Bible *je* v běžném slova smyslu neúplná[29] a snadno najdeme libovolné množství tvrzení takových, že z Bible nevyplývá ani jejich pravdivost ani nepravdivost; zjevně to však není v důsledku Gödelových vět.

Pochybenost tohoto argumentu je sice zjevná, jeho sofistikovanější podoby však lze často potkat v oblasti metodologie přírodních věd: jejich argumentace obvykle směřuje k nutné neúplnosti přírodovědných teorií v důsledku Gödelových vět.[30] Význam případné formální neúplnosti jejich matematické stránky je však zcela marginální ve srovnání s principiální nejistotou jejich správnosti danou *induktivní* metodou jejich ověřování. Neúplnost jejich matematické stránky se týká pouze jejich idealizované extrapolace do nekonečna, zatímco reálně jsou tyto teorie založeny jen na omezeném počtu pozorování a omezených výsecích reality.

Proponenti těchto interpretací rovněž často přehlížejí fakt, že ne každá teorie, která pracuje s čísly, je neúplná. Mnoho fyzikálních teorií ve skutečnosti vystačí s elementární teorií *reálných* či komplexních čísel, které jsou *úplné.* A i v těch fyzikálních teoriích, které potřebují definovat aritmetiku právě všech *přirozených* čísel, mají obvykle jakýkoli fyzikální význam přirozená čísla pouze po nějakou, třeba nepředstavitelně velkou, přesto však konečnou mez;[31] a aritmetika omezená libovolně velkým, ale *konečným* přirozeným číslem *je* (triviálně) rozhodnutelná a úplná.[32]

Právě fakt, že neúplnost se týká pouze ideálních, nekonečných matematických struktur, její aplikovatelnost v situacích týkajících se reálného světa prakticky vylučuje. Gödelovy věty tak ukazují pouze principiální nemožnost některých *idealizovaných* konstrukcí, jako byl Hilbertův program (týkající se ideální formalizace ideální matematiky) či

---

[29] Nárok na deduktivní úplnost si ostatně ani neklade.

[30] Ilustrativním příkladem je osobní výpověď fyzika Jakiho [12], jenž líčí, jak Gell-Manna přesvědčoval o nemožnosti finální fyzikální teorie v důsledku Gödelových vět.

[31] Např. mez $\alpha = 10\text{^}(10\text{^}(10^{78})))$, což nepředstavitelněkrát přesahuje nejen počet částic v pozorovatelném vesmíru, ale celkem jakýkoli rozumně myslitelný reálně existující počet; a je-li snad tato mez pro některé účely stále příliš malá, lze použít nějakou ještě větší.

[32] Neúplnost aritmetiky se týká jen tvrzení typu *Pro všechna přirozená čísla platí* ... a složitějších; nikoli např. tvrzení typu 7+5=12 či *Pro všechna* $x < 10^{78}$ *platí* ... (bez neomezených kvantifikátorů).

existence nějakých idealizovaných počítačových programů.[33] Turingovská řešitelnost je pouze velmi hrubým idealizovaným modelem reálné vyčíslitelnosti. Realitě o něco bližším modelem je zde nikoli teorie rekurze, nýbrž výpočetní složitosti, která má podstatně jemnější měřítko „praktické vyčíslitelnosti“, dané např. polynomiální složitostí. I rekurzívní, tj. principiálně algoritmicky řešitelné úlohy jsou totiž často (mají-li například exponenciální složitost) prakticky neřešitelné již pro velmi malou velikost vstupu. I tento jemnější pojem polynomiální složitosti je však jen matematická idealizace neodpovídající přesně praktické proveditelnosti (srv. například vypočítatelnost v čase omezeném polynomem stupně α pro α z pozn. 31). Praktická proveditelnost jakéhokoli výpočtu či simulace tedy záleží na mnohem přízemnějších věcech než Gödelových větách týkajících se *všech* přirozených čísel.[34]

Z přesvědčení, že *v důsledku Gödelových vět* lidská mysl přesahuje možnosti formálních systémů (a tudíž klasických výpočetních modelů) vyvozují mnozí autoři dalekosáhlé závěry. Umělá inteligence pomocí klasických počítačů je podle nich nemožná, nebo aspoň vyžaduje použití zařízení přesahujících turingovskou výpočetní sílu (např. kvantových). Činnost lidského mozku vysvětlují pomocí kvantových efektů (tak zejména Penrose, viz [17]) a často hovoří o absolutní převaze lidské mysli nad neživou přírodou (např. [15]), s nejrůznějšími světonázorovými důsledky. Toto jejich přesvědčení se zakládá na následujícím chybném argumentu:

„Gödelův důkaz ukazuje, že Gödelova sentence pro daný formální systém je v tomto systému formálně nedokazatelná, nicméně pravdivá. Avšak tím, že Gödelův důkaz chápeme, nahlížíme (my lidé) i pravdivost příslušné Gödelovy sentence – vlastně jsme ji tak dokázali mimo rámec formalizmu. Naše schopnosti (například nahlédnout pravdivost Gödelovy sentence) tak přesahují možnosti kteréhokoli formálního systému – a tedy i kteréhokoli počítače, neboť jeho činnost lze formálním systémem popsat, byť by byl sebesložitější.“

Abychom nahlédli chybnost této argumentace, stačí si uvědomit, že samu Gödelovu větu dokazujeme pomocí metod, které jsou plně formalizovatelné *v teorii množin.* Pravdivost Gödelovy sentence pro Peanovu aritmetiku (označme ji G_{PA}) tedy nahlížíme nikoli proto, že bychom nějak záhadně přesahovali možnosti počítače, nýbrž jen proto, že ji dokazujeme v silnější teorii – nikoli v PA, nýbrž v teorii množin (například ZF); tam ji však dokáže i náš počítač.

Pochopitelně v ZF zase nelze formálně dokázat Gödelovu sentenci pro ZF (G_{ZF}) – zatímco my její pravdivost (prý) dokážeme pomocí Gödelova důkazu nahlédnout. *Avšak* podle Gödelova důkazu je G_{ZF} pravdivá pouze *za předpokladu,* že je teorie ZF bezesporná. Pravdivost G_{ZF} tedy nahlížíme jen na základě víry, že ZF je bezesporná; na základě této „víry“ (tedy s přidáním axiomu Con(ZF)) však dokáže důkaz G_{ZF} zreprodukovat i náš počítač (G_{ZF} je v teorii ZF+Con(ZF) dokazatelná).

---

[33] Např. řešících zastavení *libovolně* velkých programů na *libovolně* velkých datech po *libovolně* mnoha taktech; pro programy omezené nějakou (i nepředstavitelnou) velikostí kódu α (kde α je např. číslo z pozn. 31), velikostí dat α a časem α je úloha *teoreticky* (prakticky jistě ne) rozhodnutelná prostým odsimulováním konečně (byť nepředstavitelně) mnoha kroků.

[34] To je i jedna z (většího počtu relevantních) odpovědí na častá tvrzení, že „lidská mysl nemůže být v důsledku Gödelových vět simulována počítačem“, popř. „být mechanistická“ [15], nebo že „svět nemůže být počítačovou simulací“. Mysl i svět, tak jak je vnímáme, jsou v matematickém smyslu principiálně simulovatelné už proto, že fakticky vnímáme jen konečné množství vstupů a výstupů.

Můžeme sice tento argument iterovat,[35] nikam se tím však nedostaneme. Počítač totiž v ZF umí dokázat samotnou *Gödelovu větu,* tedy i on „ví“, že Gödelova sentence *libovolného* rekurzívního formálního systému za předpokladu jeho konzistence platí. Jistotu této konzistence však v případě tak silných teorií, jako je ZF, nemá ani on, ani my.[36] Při jejím postulování nemáme v případě dostatečně silné teorie o nic větší jistotu než počítač, který ji přijme mechanicky; naše „neformální nahlédnutí“ pravdivosti Gödelovy sentence G_T není tedy o nic lepší, než její formální důkaz v T+Con(T).[37]

Je tedy sice možné (byť sotva pravděpodobné), že schopnosti lidské mysli jsou dány kvantovými efekty v mikrotubulech neuronů (jak tvrdí Penrose); rozhodně to však nijak nevyplývá z Gödelových vět a ani to jimi nemůže být heuristicky podloženo.[38] Ze všech těchto důvodů lze konstatovat, že Gödelovy věty nedávají žádnou indicii, že by projekt umělé inteligence byl nemožný, a to ani v rámci stávajících (turingovských) výpočetních modelů (na „klasickém“ počítači).

Poznámka: Práce na článku byla částečně podpořena grantem č. GD 401/03/H047 Grantové agentury České republiky.

Literatura

Reference na online zdroje se vztahují k lednu 2007. Některé z těchto zdrojů vyžadují předplatné (prola.aps.org, jstor.org, springerlink.com).

[1] Balcar, B., Štěpánek, P.: *Teorie množin.* Praha: Academia 2001 (1. vydání 1986).
[2] Feferman, S.: Gödel's program for new axioms: Why, where, how and what? In P. Hájek (ed.): *Gödel '96.* New York: Springer 1996: 3–22. (Preprint na math.stanford.edu/~feferman)
[3] Goldsteinová, R.: *Neúplnost. Důkaz a paradox Kurta Gödela.* Dokořán, Praha 2006.
[4] Gödel, K.: Die Vollständigkeit der Axiome des logischen Funktionenkalküls. *Mh. Math. Phys.* **37** (1930) 349–360.
[5] Gödel, K.: Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I. *Mh. Math. Phys.* **38** (1931) 173–198. (Částečný anglický překlad na www.research.ibm.com/people/h/hirzel)
[6] Gödel, K.: Zum intuitionistischen Aussagenkalkül. *Anz. Akad. Wiss. Wien* **69** (1932) 65–66.

[35] Počítač neumí dokázat $G_{ZF+Con(ZF)}$, avšak ani my ji „nenahlížíme“ bez víry v Con(ZF+Con(ZF)), s níž ji dokáže i počítač atd.

[36] Naše přesvědčení o bezespornosti ZF je pouze empirické – za sto let se v ní žádný spor nenašel. Lze sice tvrdit, že vnitřním zrakem si (na rozdíl od počítače) univerzum množin umíme „představit“. Historické zkušenosti s postulováním sporných teorií (Cantorova naivní teorie množin, Fregova logika, původní Churchova verze λ-kalkulu aj.) však ukazují, že toto přesvědčení může být zcela mylné.

[37] Pro obdobnou formulaci tohoto vysvětlení mylnosti původního argumentu viz též [18].

[38] Někteří autoři [15] ve snaze zachránit svůj závěr sklouznou do logických kruhů, rozmlžených formulací a neadekvátních popisů (mysl a jen mysl může „z principu“ tvrdit svoji konzistenci apod.). Nakonec mohou pouze uzavřít, že mysl není *jednoduchá* mašinka [15], což je zřejmé i bez Gödelových vět.

[7] Gödel, K.: The consistency of the axiom of choice and of the generalized continuum-hypothesis. *Proc. Nat. Acad. Sci. USA* **24** (1938) 556–557. (Online na www.pubmedcentral.gov)
[8] Gödel, K.: An example of a new type of cosmological solutions of Einstein's field equations of gravitation. *Rev. Mod. Phys.* **21** (1949) 447–450. (Online na prola.aps.org)
[9] Gödel, K.: *Filozofické eseje.* Praha: Οικουμενη 1999.
[10] Hájek, P.: A new small emendation of Gödel's ontological proof. *Studia Logica* **71** (2002) 149–164. (Online na www.springerlink.com)
[11] Hofstadter, D. R.: *Gödel, Escher, Bach: An Eternal Golden Braid.* New York: Basic Books 1979.
[12] Jaki, S. L.: A late awakening to Gödel in physics. *Sensus Communis* **5** (2004) No. 2–3. (Online na pirate.shu.edu/~jakistan)
[13] Kleene, S. C.: The work of Kurt Gödel. *J. Symb. Log.* **41** (1976) 761–778. (Online na www.jstor.org)
[14] Kreisel, G.: Kurt Gödel. 28 April 1906 – 14 January 1978. *Biographical Memoirs of Fellows of the Royal Society.* The Royal Society 1980. (Online na www.jstor.org)
[15] Lucas, J. R.: Minds, machines and Gödel. *Philosophy* **36** (1961) 112–127. (Online na users.ox.ac.uk/~jrlucas)
[16] Nagel E., Newman, J. R.: *Gödelův důkaz.* Brno: Vutium 2003.
[17] Penrose, R.: *Makrosvět, mikrosvět a lidská mysl.* Praha: Mladá fronta 1999.
[18] Podnieks, K.: *Вокруг теоремы Геделя.* Riga: Zinatne 1992. (Ruský originál a anglický překlad *What Is Mathematics: Gödel's Theorem and Around* na www.ltn.lv/~podnieks)
[19] Smullyan, R.: *Jak se jmenuje tahle knížka?* Praha: Mladá fronta 1986.
[20] Smullyan, R.: *Navěky nerozhodnuto.* Praha: Academia 2003.
[21] Sochor, A.: *Klasická matematická logika.* Praha: Karolinum 2001.
[22] Sochor, A.: *Metamatematika teorií množin.* Praha: Karolinum 2005.
[23] Švejdar, V.: *Logika. Neúplnost, složitost a nutnost.* Praha: Academia 2002.
[24] Turing, A. M.: On computable numbers, with an application to the Entscheidungsproblem, *Proceedings of the London Mathematical Society* 2/**42** (1936) 230–265. (Online na www.turingarchive.org)
[25] *Wikipedia,* en.wikipedia.org, heslo „Kurt Gödel" a hesla jím odkazovaná.
[26] Willard, D. E.: Self-verifying axiom systems, the incompleteness theorem and related reflection principles. *J. Symb. Log.* **66** (2001) 536–596. (Online na www.jstor.org)
[27] Zlatoš, P.: *Ani matematika si nemôže byť istá sama sebou.* Bratislava: Iris 1995 (Online verze na thales.doa.fmph.uniba.sk/zlatos)
[28] Zlatoš, P.: Gödelov ontologický dôkaz existencie Boha. *Organon F* **3** (1996) 211–238.

Neurovedné okno do vedomia

Ľubica Beňušková[1]

Abstrakt. Názor, že vedomie sprevádza aktivitu vysoko a špeciálne organizovanej hmoty mozgu, nadobudol v súčasnosti detailnejšie argumenty. Je možné, že tak ako sa na určitom stupni vývoja a organizácie hmoty objavil život, tak sa na určitom stupni vývoja a organizácie *živej* mozgovej hmoty objavuje vedomie, hoci stále nevieme povedať ako a prečo.

1 Úvod

Túžba stvoriť umelú bytosť z neživých súčiastok a vdýchnuť jej život je veľmi stará (Golem, Frankenstein, Sirael magistra Kellyho). V súčasnosti existujú vedecké a technické disciplíny, ako je umelá inteligencia, umelý život, robotika, atď., ktoré túto túžbu postupne realizujú. Budú mať niekedy v budúcnosti tieto roboty, tieto bytosti vedomie, uvedomia si sami seba? Aby sme vedeli odpovedať na túto otázku „áno“ alebo „nie“, museli by sme najskôr vedieť, čo to vedomie vlastne je. V súčasnosti mnohí vedci i filozofi považujú problém vedomia za vyriešený. Problémom je, že týchto riešení je viacero. Podľa Searleho je vedomie ekvivalentné špecifickým procesom prebiehajúcim iba v mozgu [29]. Dennett tvrdí, že vedomie a ďalšie mentálne procesy sú ekvivalentné špecifickým výpočtom [5]. Podľa Edelmana je kľúčom k vedomiu nesmierna zložitosť mozgu [6]. Kováč tvrdí, že vedomie ako organizovaný súhrn kválií[2], ktoré sa objavilo ako fenomenologický výsledok anatomickej komplexifikácie mozgu, je produktom evolučnej eskalácie emocionálneho výberu [14]. Ktorý názor je správny? Sú správne všetky, pričom každý z nich kladie dôraz len na vybrané aspekty tohto javu? Je to naozaj také samozrejmé a prirodzené, že vedomie sa objavuje, keď sú splnené isté podmienky? Aké podmienky sú to?

V jeho vplyvnej knihe *Shadows of the Mind* fyzik Roger Penrose argumentuje, že problém vysvetlenia vedomia patrí do domény elementárnych zákonov fyziky [22]. Podľa Penrosa, v tomto fyzikálne determinovanom svete zatiaľ neexistuje žiadna teória, a to ani fyzikálna teória, ktorá by nám pomáhala vysvetliť záhadu psychična vrátane vedomia. Penrose volá po radikálnej zmene v samom základe ponímania elementárnych zákonov fyziky. Vidí cestu v úplne novej teórii spájajúcej Einsteinovu všeobecnú teóriu relativity s kvantovou teóriou. Túto novú teóriu však ostáva ešte objaviť. Miestom, či nositeľom koherentných kvantových stavov zodpovedajúcich vedomiu by mali byť mikrotubule, filamentárne štruktúry vo vnútri neurónov. Táto nová teória si musí poradiť s mnohými problémami vrátane kvantového kódovania informácie v mikrotubulách alebo iných relevantných nosičoch prenosu kvantovej informácie do pálenia neurónov, ktoré predchádza vôľou riadenému pohybu svalov, udržiavania koherentných stavov na makroskopické vzdialenosti po dobu stovák milisekúnd, apod. [13]. Penrosovi kritici spochybňujú, či fyzika vôbec môže niečo dôležité kompetentne povedať o psychických

[1] Katedra aplikovanej informatiky, Fakulta matematiky, fyziky a informatiky, Univerzita Komenského, Mlynská dolina, 842 48 Bratislava 4, E-mail: benus@ii.fmph.uniba.sk.

[2] Filozofi používajú termín kváliá (qualia) na označenie tých aspektov nášho psychického života, ktoré sú prístupné iba introspekcii (introspectively accessible, phenomenal aspects of our mental lives) [33].

javoch vo všeobecnosti a konkrétne o vedomí (pozri http://psyche.cs.monash.edu.au/psyche-index-v2.html#som).

V inej vplyvnej knihe na tému vedomia, filozof a matematik, David J. Chalmers jasne argumentuje, že vedomie a psychika sú skutočnou záhadou, hlavne čo sa týka tzv. ťažkého problému kválií [3]. Chalmers sa pýta prečo by kvantové procesy alebo akékoľvek iné špecifické fyzikálne procesy v mikrotubulách, neurónoch, či iných mozgových štruktúrach, alebo akékoľvek výpočtové procesy mali dať vznik vedomiu (rozumej kváliám)? Ak chceme brať vedomie (kváliá) vážne, Chalmers vraví, musíme prekročiť striktne materialistický rámec. Podľa neho však fundamentálne zákony spájajúce fyzické a psychické ostáva ešte objaviť.

V tejto kapitole budeme pod pojmom vedomie chápať „nedefinovateľný produkt niečoho, čo je veľmi organizované, samo nie je vedomé, porovnateľné so svetlom“ [34]. Pod týmto nevedomým vysoko organizovaným niečím budeme chápať mozog. Preberieme špecifické procesy, ktoré prebiehajú v mozgu, a ktoré majú nejaký súvis s fenoménom vedomia za cieľom hlbšieho pochopenia záhadného spojenia medzi hmotou a vedomím.

2 Vedomie nejako súvisí so špecifickou aktivitou mozgu

Aké sú vlastne indície, že vedomie špecificky súvisí s aktivitou mozgu? Najskôr by sme chceli spomenúť zaujímavý časový aspekt. Známe experimenty Benjamina Libeta na ľudských pacientoch podstupujúcich operáciu mozgu pri plnom vedomí ukázali, že to trvá asi pol sekundy (500ms), aby sa objavilo uvedomenie si rozhodnutia konať [16]. Všetky pohyby svalov inicializované slobodnou vôľou sú predchádzané merateľnou špecifickou elektrickou zmenou v mozgu, ktorá sa volá pohotovostný potenciál (readiness potential, RP). RP sa objavuje v mozgu v priemere 550ms pred aktom samotným. *Ľudia si uvedomili svoje vlastné rozhodnutie konať 350 až 400ms po objavení sa RP* a asi 200ms pred vykonaním aktu. Tých 200ms zodpovedá času nevyhnutnému na iniciáciu a generovanie pohybov svalov. Tieto dáta vyvolávajú veľa zásadných otázok (čo sa týka detailov experimentov odporúčame čitateľa na pôvodné práce Libeta). Napríklad, otázky o podstate slobodnej vôle, či vôbec môžeme o takom niečom hovoriť. Zatiaľ čo Libet ostáva v tejto záležitosti optimistom [17], otázka prečo to trvá tak dlho, pokiaľ vznikne uvedomenie, stále nie je zodpovedaná. Prečo vôbec uvedomenie vznikne? Ktoré procesy sú za to zodpovedné a ako k tomu dochádza?

Ďalšou indíciou o súvise psychiky s mozgom sú napríklad tzv. psychotropné látky, ako sú drogy a lieky na duševné poruchy. Tieto látky sú psychotropné preto, lebo majú chemickú štruktúru podobnú na prirodzené neurotransmitery, a preto môžu chemicky reagovať s ich synaptickými receptormi. Napríklad jedna skupina liekov znižujúcich napätie a úzkosť, tzv. anxiolytiká, pôsobí na inhibičné GABA-ergné receptory. Antidepresíva, lieky proti depresii, pôsobia na serotonínové a/alebo noradrenalínové receptory a upravujú tak hladiny týchto látok v mozgu. Lieky podávané pri schizofrénii a iných psychotických stavoch pôsobia na dopamínové receptory. Nadmerne znížená, resp. zvýšená aktivita neurotransmiterových systémov vedie k rôznym poruchám psychických funkcií [9,10]. K poruchám v hladine neurotransmiterov môže dôjsť napríklad v dôsledku vrodených genetických predispozícií alebo v dôsledku drogovej závislosti alebo následkom iných faktorov. Totižto, produkcia neurotransmiterov, receptorov a vôbec všetkých molekúl v neurónoch je naprogramovaná geneticky. Na tento

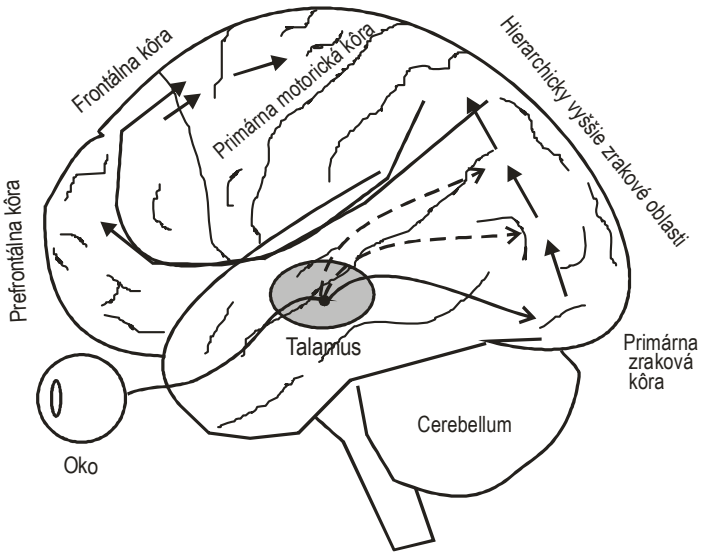
vrodený genetický program vplýva stres, vek, iné látky ako napríklad hormóny. Manipulácie s chemizmom mozgu majú prechodné alebo trvalé psychické následky [10].

Ďalšou oblasťou súvisu psychického prežívania s anatómiou a fyziológiou mozgu sú javy z oblasti somestézie. V jednej štúdii sa pomocou magnetoencefalografie skúmal súvis mozgových reprezentácií prstov pred operáciou a po operácii syndaktýlie (vrodeného zrastenia prstov na ruke) so subjektívnym vnímaním [21]. Pred operáciou títo pacienti vnímali, cítili tieto prsty ako jeden celok. Ešte aj niekoľko týždňov po ich chirurgickom oddelení. Subjektívnu zmenu vo vnímaní svojich prstov sprevádzala zmena nesomatotopickej reprezentácie ruky na somatotopickú v dôsledku oddeleného používania prstov a mozgovej plasticity. Poukazuje to na to, že existencia mozgových máp (reprezentácií) priamo súvisí s našimi pocitmi. U 95 až 100 percent ľudí, ktorí stratili rameno alebo nohu, sa prejavuje fenomén fantómovej končatiny (phantom limb) [20]. Cítia svoju končatinu, cítia na nej dotyky, tlak na kožu, teplo, chlad, únavu, bolesti. Cítia ju, akoby ju stále mali, cítia, ako sa hýbe, keď sa oni hýbu, je stále súčasťou ich subjektívnej skúsenosti seba samých. Fantómová končatina má spočiatku normálnu veľkosť a tvar, no časom sa začína meniť: môže sa ocitnúť voľne vo vzduchu, scvrknúť sa do kýpťa, alebo nadobudnúť iný bizarný tvar či veľkosť alebo spojenie s telom. Stále je však integrálnou súčasťou pacientovho ja.

Mohli by sme takto pokračovať ďalej. V súčasnosti už neurovedné a psychologické pozorovania a experimenty dovoľujú istú konkrétnu predstavu o tzv. neurálnych korelátoch vedomia. Súčasná neuroveda má dobré dôvody predpokladať, že neurálne koreláty reflektívneho vedomia sú odvodené od neurálnych korelátov vnímania, teda že reflektívne vedomie je vybudované na princípoch primárneho respektíve predreflektívneho vedomia. *Predreflektívne vedomie* je svet zmyslových (zrakových, sluchových, somestetických, atď.) vnemov. *Reflektívne vedomie* je to, keď ja si uvedomujem, že Vy si uvedomujete, že druhí si uvedomujú, že ja si uvedomujem. Najprv sa budeme venovať neurálnym korelátom zrakového uvedomovania (visual awareness), ako v súčasnosti najviac preštudovaného zmyslového systému.

3 Na uvedomenie je potrebná nadprahová kvantita neurónov

Veľa sa o mechanizmoch zrakového uvedomovania zistilo pri štúdiu mechanizmov tzv. slepého videnia (blindsight) [27]. Keď sa poškodí celá primárna zraková kôra (pozri Obr. 1), pričom zraková dráha od oka po kôru ostane nepoškodená, človek oslepne, nevidí. Nazýva sa to kortikálna alebo kôrová slepota. Kortikálna slepota sa líši od slepoty, ktorá je spôsobená poškodením oka, sietnice alebo zrakového nervu. Tieto poškodenia totiž eliminujú vstup zrakových podnetov do mozgu, zatiaľ čo deštrukcia primárnej zrakovej kôry nie. Ukážme si to na Obr. 1: podnety z oka idú cez zrakové jadro v talame do primárnej zrakovej kôry. Stadiaľ idú do hierarchicky vyšších oblastí zrakovej kôry. Existujú aj riedke spojenia z talamu do vyšších zrakových oblastí, ktoré obchádzajú primárnu zrakovú kôru (na Obr. 1 sú znázornené prerušovanou čiarou) a existujú aj spojenia z oka do podkôrových štruktúr (ktoré nie sú na obrázku znázornené).



Obrázok 1. Ľudský mozog. Povrch mozgu tvorí silne zvrásnená mozgová kôra, sídlo kognície. Evolučne najmladšou je prefrontálna kôra. Podkôrové časti (ako napr. talamus) sú evolučne staršie ako kôra. Cez talamus idú všetky zmyslové podnety. V mozgu človeka sa nachádza rádovo 10^{11} nervových buniek - neurónov. Každá z nich vytvára kontakty na rádovo 10-tich tisícoch iných neurónov a na sebe prijíma signály z kontaktov od iných rádovo 10-tich tisícov neurónov. Rýchle spracovanie signálov v rámci jednej bunky má elektrickú povahu a prenos signálov z jednej bunky na druhú má chemickú povahu. Jednotlivé funkčné oblasti tvoria navzájom recipročne husto pospájané a hierarchicky zorganizované neurónové siete. Napríklad v zrakovej kôre sa uvažuje asi o 20 hierarchicky rôznych oblastiach. Obrázok je prevzatý z [1].

K poškodeniu primárnej zrakovej kôry môže dôjsť napr. následkom úrazu, náhlej mozgovej príhody, nádoru, apod. Na prekvapenie vedcov, ktorí vyšetrovali takýchto pacientov, sa zistilo, že ostávajúce časti zrakového systému majú obmedzenú schopnosť spracovať vizuálne podnety. Niektorí takto slepí pacienti (nie všetci) dokážu využiť zrakové podnety na identifikáciu polohy, pohybu a smeru pohybu objektu. Sú dokonca schopní rozlíšiť jednoduché formy, a to tak, že primerane utvarujú svoju ruku, keď sú požiadaní ako by daný objekt uchopili. Toto všetko bez toho, aby boli schopní daný podnet vedome vnímať, a teda vidieť v bežnom slova zmysle. Správnosť vykonania úlohy sa v týchto experimentoch vyhodnocuje štatisticky, to znamená, že za slepé videnie sa považuje úspech vo viac pokusoch ako by zodpovedalo náhode. Pacient si nie je vedomý týchto svojich schopností a experimentátori ho musia pobádať k tomu, aby sa pokúsil niečo hádať. Napríklad ho žiadajú, aby ukázali prstom na objekt. Pre slepého pacienta je to nezmyselná úloha, lebo objekt nevidí. Podľa svojej interpretácie iba háda a robí niečo len preto, aby vyhovel experimentátorovi. Ale aj tak ukáže správnym smerom štatisticky signifikantne viac krát ako by zodpovedalo náhode. Na základe zistenej zvýšenej aktivity neurónov v zostávajúcom zrakovom systéme sa usudzuje, že tie riedke spojenia, ktoré obchádzajú poškodenú primárnu zrakovú kôru sú schopné viesť k pohybovej odpovedi pacienta [28].

Hypoteticky usudzujeme, že tieto dráhy nestačia na vznik vedomého vnemu a to preto, lebo nie je aktívne dostatočné množstvo (kvantita) neurónov. Požiadavka aktivácie nadprahovej kvantity neurónov sa prvýkrát objavila v úvahách Francisa Cricka a Christopha Kocha [4, 12].

4 Synchrónne oscilácie sprevádzajú vnímanie

Binokulárna rivalita (binocular rivalry) vzniká, keď každému oku premietame iný obraz. Zrakový vnem, ktorý vnímame, nie je ani priestorovou sumou ani priemerom týchto dvoch obrazov. Dochádza k striedaniu týchto obrazov (vnemov), ako keby navzájom súťažili. Tento jav sa študuje aj na opiciach, u ktorých sa pomocou elektród priamo meria aktivita kôrových neurónov. Logothetis a jeho spolupracovníci [15] objavili, že v rôznych hierarchických častiach zrakovej kôry existujú neuróny, ktorých aktivita je korelovaná buď s jedným alebo druhým vnemom. V podobnej štúdii na mačkách, Fries et al. [7] objavili, že tieto neuróny navzájom krátkodobo, rádovo 100-ky ms synchronizovane oscilujú. Teda, že ich výboje sa periodicky synchronizujú. Neuróny totižto vydávajú elektrické impulzy. V skupine neurónov môžu jednotlivci vydávať impulzy kedy chcú, v náhodnom poradí, alebo istá podskupina môže vydávať impulzy naraz, t.j. synchronizovane.

Analogické procesy boli zistené aj u ľudí. Experimentálne pomocou analýzy EEG sa zistilo, že zrakové uvedomenie si tváre na základe vysokokontrastnej fotografie bol u ľudí sprevádzaný takouto krátko trvajúcou (180ms) synchronizáciou aktivity neurónov medzi hierarchicky najvyššou zrakovou oblasťou v posteriórnej parietálnej kôre a frontálnou kôrou (pozri Obr. 1) [27]. Je zaujímavé, že táto synchronizácia nastala iba v ľavej hemisfére, ktorá je tzv. vedomou hemisférou [9]. U väčšiny ľudí tam sídli spracovanie reči. Pri pozeraní na nezmyselné škvrny, ktoré vznikli po otočení vysokokontrastnej fotografie tváre o 180°, takáto synchronizácia ale ani vnem tváre nenastali. K synchronizácii aktivity došlo neskôr aj v pohybovej (motorickej) kôre oboch hemisfér. Táto synchronizácia sprevádzala pohybovú odpoveď subjektu – t.j. stlačenie tlačítka indikujúceho, či vidí tvár alebo „netvár“. Prvý bod hypotézy zrakového uvedomenia je tento:

- V hierarchicky najvyšších oblastiach mozgovej kôry, ako parietálna (temenná) zraková kôra – frontálna kôra, musí byť po istý čas (rádovo stovky milisekúnd, 100-200ms) synchrónne aktívna istá kvantita, isté nadprahové množstvo neurónov.

5 Časti mozgu zodpovedné za plánovanie „riadia“ aj vnímanie

Na psychologickej úrovni, uvedomovanie si senzorického objektu zahŕňa upriamenie pozornosti na tento objekt. Kôrové oblasti aktívne počas upriamovania pozornosti sa nachádzajú vo frontálnom a parietálnom laloku, v ľavej i pravej hemisfére. Ich aktivácia sprevádza subjektívne vedomé prežívanie a vedomé riešenie kognitívnych úloh [24]. V parietálnom laloku sa nachádzajú hierarchicky najvyššie zrakové oblasti. Iné časti frontálnej a prefrontálnej kôry sú aktívne pred tým, ako vykonávame vôľou riadenú činnosť. Ich poškodenie vedie k poškodeniu plánovania vôľou riadených pohybov a správania. Čím hierarchicky vyššie je daná frontálna oblasť (teda čím je ďalej vpredu, pozri Obr. 1), tým je časová škála jej plánovania našich pohybových aktivít upriamená na dlhšiu dobu dopredu. Primárna motorická kôra už riadi priamo svaly. Jej činnosti časovo predchádza činnosť plánovacích oblastí kôry. Frontálna a prefrontálna kôra sú evolučne najmladšie a svojou plochou sú najväčšie u človeka. Sídlia v nich plánovacie a pozornostné funkcie. Plánovacie a pozornostné oblasti vo frontálnom laloku sa síce prekrývajú, ale nie sú totožné. Pozornosť môže byť tiež riadená vôľou, čiže plánovacie

oblasti sú hierarchicky najvyššie. Prefrontálna, plánovacia časť mozgovej kôry sa vyvíja po narodení a jej neurónové siete sa organizujú vďaka učeniu.

Medzi plánovacími a pozornostnými prednými časťami mozgovej kôry a „zadnými“ senzorickými, dochádza k „bottom-up“ a „top-down“ interakciám. Ak dôjde na senzorickej úrovni k nadprahovému zorganizovaniu synchrónne oscilujúcich neurónov, napríklad pri silnom podnete, upriami sa naň pozornosť. Spontánna organizácia na senzorickej úrovni tak môže slúžiť aj na manažovanie pozornosti („botttom-up“). Zase naopak, upriamovanie pozornosti, čiže aktivácia pozornostných oblastí frontálnej kôry, môže vďaka spätnoväzbovým spojeniam do parietálnej kôry ovplyvniť, ktoré senzorické neuróny sa synchronizujú („top-down“) [30]. Leopold a Logothetis [15] tiež predpokladajú, že aktivita plánovacích centier neustále „mieša karty“ v senzorických oblastiach. Vedie ich k tomu zistenie, že hoci striedanie vnemov pri binokulárnej rivalite (a iných javoch multistabilného vnímania) je spočiatku náhodné a nevedomé, je možné sa naučiť ho vôľou ovládať. Rozšírime predchádzajúci bod hypotézy vedomého vnímania takto:

- Synchronizovaná aktivácia nadprahovej kvantity neurónov v hierarchicky najvyšších senzorických[3] oblastiach musí synchrónne spolupracovať s aktiváciou centier súvisiacich s našou pozornosťou a vôľovou činnosťou**.**

6 Komunikácia je na vznik reflektívneho vedomia nevyhnutná

V predchádzajúcich častiach sme si predstavili hypotetické neurálne koreláty zrakového uvedomenia, respektíve predreflektívneho vedomia. Wolf Singer (spoluobjaviteľ synchrónnych oscilácií súvisiacich s vnímaním [7]) si myslí, že reflektívne vedomie by vzniklo z takých istých procesov, ale tieto procesy by sa neaplikovali na signály prichádzajúce zo zmyslových častí kôry, lež na signály pochádzajúce z asociačných častí mozgovej kôry, ktoré zahŕňajú hlavne oblasti spracúvajúce jazyk (reč a text) a oblasti, v ktorých sídlia naše predstavy, pamäť a emócie [31]. Čiže reflektívne vedomie by zodpovedalo akémusi „zraku mysle, ktorý vníma reprezentácie nášho vnútorného sveta“. Avšak k neuronálnej hypotéze reflektívneho vedomia pridáva ešte ďalší bod, a to požiadavku komunikácie medzi mozgami, potenciálnymi nositeľmi vedomia. Na uvedomenie si toho, že Vy si uvedomujete, že ja si uvedomujem, musia byť komunikujúce mozgy schopné vytvoriť si model stavu toho druhého mozgu a toto nie je bez komunikácie možné.

Sebareflexia je možná vďaka komunikácii a sociálnej interakcii. Lenže zdá sa, že mozog má na tieto reprezentácie vyhradené pripravené štruktúry. V roku 1996 rozvíril diskusiu vo vedeckých kruhoch objav, ktorý urobila skupina Taliana G. Rizzolattiho [26]. Snímali aktivitu neurónov vo ventrálnej premotorickej frontálnej kôre opíc a zistili, že pri rôznych špecifických činnostiach sú aktívne rôzne neuróny. Jednou z činností bolo napríklad uchopenie orieška a vloženie si ho do úst, ďalšou ošúpanie banánu a pod. To by nebolo až také prekvapujúce, keďže tieto neuróny majú na starosti prípravu pohybov. Šokujúce však bolo zistenie, že tie isté neuróny boli rovnako aktívne aj vtedy, keď opica sedela nepohnute a *pozorovala* iných, druhú opicu alebo experimentátora pri vykonávaní tej istej činnosti. Rizzolatti nazval tieto neuróny zrkadliace neuróny (mirror neurons). Čiže kedykoľvek pozorujeme niekoho niečo robiť, príslušné zrkadliace neuróny v našej

[3] Všetky úvahy týkajúce sa zrakového systému platia aj na ostatné zmyslové systémy.

premotorickej frontálnej kôre „imitujú“ danú činnosť, a umožňujú nám tak pochopiť a porozumieť aktivite druhých bytostí. Zrkadliace neuróny môžu tvoriť neurálnu bázu pre také enigmatické aspekty ľudskej mysle, ako sú učenie napodobovaním, „čítanie mysle“, respektíve úmyslov druhých, empatia a pod. Ešte zarážajúcejšie je, že u ľudí sa vo ventrálnej premotorickej kôre nachádza tzv. Brocova oblasť, ktorá je mozgovým centrom asociovaným s expresívnymi (vokalizačnými) a syntaktickými aspektami jazyka. Čiže schopnosť zrkadliť intencie iných ľudí a schopnosť napodobovať ich vokalizáciu sídlia u ľudí v rovnakej oblasti mozgu. Možno práve táto ventrálna premotorická kôra a jej vývoj v evolúcii zohrali determinujúcu úlohu pri vzniku jazyka.

Normálni ľudia majú schopnosť vysvetliť a predikovať správanie iných ľudí na základe predpokladania ich myšlienok a pocitov. Táto schopnosť prisudzovať mentálne stavy iným, ale aj *sebe*, sa nazýva „mentalizácia“ alebo „teória mysle“ (theory of mind). Neinvazívne zobrazovanie mozgovej aktivity ukázalo, že schopnosť pripisovať rozličné mentálne stavy, ako predstavy a želania iným aj sebe, závisí od dobrého fungovania špecifického neurokognitívneho modulu, ktorý zahŕňa: (i) mediálnu prefrontálnu kôru, konkrétne anteriórnu časť paracingulárnej kôry, (ii) časť horného temporálneho závitu a (iii) temporálnu kôru susediacu s amygdalou [8]. Asi nie je náhoda, že tento modul prechádza cez plánovacie a emočné časti mozgu. Aktivita neurónov v tomto neurokognitívnom module teórie mysle u ľudí trpiacich autizmom je podstatne nižšia – navonok sa to prejavuje akousi „mentálnou slepotou“ (mind blindness) [8]. Treba pripomenúť, že existuje celé spektrum autistických porúch, od miernych po mimoriadne závažné, teda ani tu neplatí „všetko, alebo nič“.

Na možný súvis reflektívneho vedomia s učením[4] a komunikáciou poukazuje i to, že vedomie sa v priebehu života človeka *vyvíja*, a to smerom od primárneho pred-reflektívneho vedomia k reflektívnemu. Úroveň vedomia trojročného dieťaťa je iná ako úroveň (či reflektívnosť) 12-ročného alebo 42-ročného človeka.

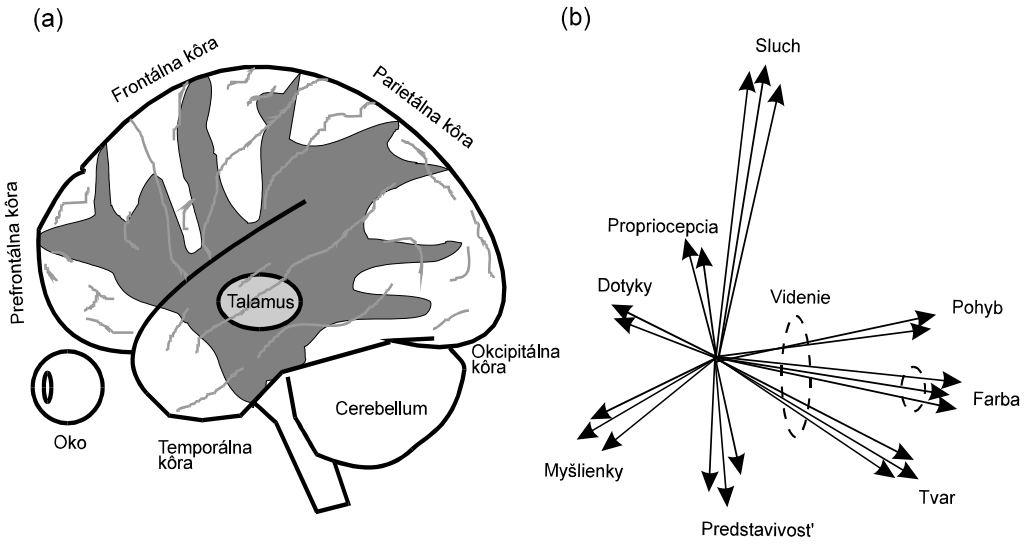
7 Dynamické jadro: oheň produkujúci „svetlo“?

Pomocou súčasných techník zaznamenávania elektromagnetickej aktivity mozgu, ako je počítačová magnetoencefalografia (MEG), je možné zachytiť, analyzovať a zobraziť rýchlu semiglobálnu koherentnú aktivitu v mozgu, ktorá sprevádza vedomú percepciu senzorického podnetu [32]. Je to takmer magický obraz, pretože táto semiglobálna koherentná aktivita sa mení v čase ako horiaci oheň s centrom v talame, ktorého „plamene“ zasahujú a pohlcujú práve zosynchronizované oblasti mozgu (pozri Obr. 2). G. Edelman and G. Tononi [6] nazývajú túto premenlivú semiglobálnu koherentnú aktivitu *dynamické jadro*. Prečo je plameň živený z talamu? Klinický výskum ukázal, že poškodenie intralaminárnych jadier v talame vedie ku strate vedomia a ku kóme. Neuróny v intralaminárnych talamických jadrách majú husté recipročné spojenia s mozgovou kôrou. Intralaminárne jadrá sú zdrojom vybudenia (arousal), bez ktorého kôra nemôže fungovať. Navyše talamus spolugeneruje talamokortikálne gama oscilácie, ktoré sa generujú a udržiavajú v spätnoväzbovej interakcii medzi kôrou a talamom počas kognitívnych aktivít [25]. Dynamické jadro zodpovedá veľkému (semiglobálnemu) spojitému klastru populácií neurónov, ktoré sú koherentne aktívne na časovej škále stovky

[5] V súčasnosti sa vie, že učenie je v mozgových neurónových sieťach sprevádzané zmenami váh (účinností) medzineurónových spojení – synáps. Dôsledkom je zmena výstupu siete vyvolaná daným vstupom.

milisekúnd. Participujúce neurónové populácie sú počas tejto doby oveľa viac prepojené medzi sebou než s ostatnými časťami mozgu. Táto prepojenosť umožňuje rýchlu a efektívnu komunikáciu medzi participujúcimi neurónovými sieťami. V nedávnej štúdii bolo ukázané, že absencia vedomia pri aktívnom zdravom mozgu počas istých fáz spánku je sprevádzaná kolapsom efektívnej kortikálnej konektivity [19].

Konkrétna kompozícia dynamického jadra sa mení zhruba každých 150ms. Každých asi 150ms vznikne nový vzor semiglobálnej aktivity. Existuje nekonečné množstvo možností, podobne ako v prípade plameňov ohňa. Ako ukázala počítačová MEG, tak ako akékoľvek dve kompozície dynamického jadra nikdy nie sú rovnaké v mozgu jedného človeka, tak nie sú rovnaké ani medzi dvoma rôznymi ľuďmi.



Obrázok 3. (a) Ilustrácia dynamického jadra. Dynamické jadro je premenlivá semiglobálna koherentná aktivita mozgu. Predpokladá sa o nej, že je neuronálnym korelátom vedomia. „Plamene“ jadra tancujú ako plamene ohňa. Jedna konfigurácia jadra trvá asi 150ms. (b) Interpretácia dynamického jadra ako N-dimenzionálneho neuronálneho referenčného priestoru, kde každá os (dimenzia) označuje nejakú populáciu neurónov, ktorá kóduje (reprezentuje) daný aspekt vedomej skúsenosti. Každá os na obrázku sa dá rozložiť ešte na elementárnejšie osi. V skutočnosti sú stovky tisíc takýchto dimenzií. Obrázok je prevzatý z [2].

Podľa Edelmana a Tononiho [6], dynamické jadro pozostáva z veľkého počtu distribuovaných skupín neurónov, ktoré sú prechodne súčasťou jadra, keď sú navzájom koherentne aktívne. Neuronálny referenčný priestor pre ľubovoľný stav vedomia sa dá ilustrovať ako abstraktný N-dimenzionálny priestor, stavový priestor obrovskej neurónovej siete mozgu, kde každá os (dimenzia) reprezentuje nejakú konkrétnu populáciu neurónov, ktorá kóduje daný aspekt vedomej skúsenosti. Vzdialenosť na osi od počiatku súradnicovej sústavy reprezentuje výraznosť daného aspektu vedomej skúsenosti. Táto výraznosť môže byť napríklad kódovaná množstvom neurónov, ktoré sú v danej populácii práve aktívne, či všetky (hodnota 1 na osi), polovica (hodnota ½ na osi), atď. Kľúčovým pojmom subjektívnej skúsenosti sú kváliá. Kváliá sú špecifické kvality subjektívnej skúsenosti, ako napríklad subjektívne vnímanie červenosti, príjemného tepla, bolesti zubov, atď. Podľa teórie dynamického jadra čistá červenosť by bola

reprezentovaná len jediným konkrétnym stavom dynamického jadra, to jest jediným bodom v N-dimenzionálnom neuronálnom stavovom priestore [6]. V tomto stave jadra by hlavne participovala populácia neurónov, ktoré kódujú červenú farbu a ostatné populácie neurónov v žiadnej alebo veľmi malej miere. Koordináty bodu v N-rozmernom neuronálnom referenčnom priestore sú determinované aktivitami všetkých neurónových populácií, ktoré sú súčasťou jadra (ostatné koordináty sú rovné nule). Samozrejme všetky koordináty sa v čase menia. Teda subjektívny zážitok tej istej červenej bude rôzny u rôznych ľudí a môže sa líšiť aj u tej istej osoby napríklad ráno a večer. Celistvosť dynamického jadra zodpovedá psychickej celistvosti obsahu vedomia.

Čo by bola v tejto teórii neuronálna báza podvedomia? Tá istá populácia neurónov môže byť v jednom čase súčasťou koherentnej aktivity jadra a reprezentovať časť vedomej skúsenosti a inokedy môže vykonávať svoje spracovanie oddelene a byť tak zahrnutá v nie vedomom spracovaní. Koch a Crick [11] navrhli, že tie neurónové siete, ktoré sa momentálne nezúčastňujú na semiglobálnej koherentnej aktivite, môžu stále spracovávať svoje vstupy. Výsledky ich spracovania v konečnom dôsledku skôr alebo neskôr ovplyvnia myslenie a správanie, čo však asi nebudeme vedieť vysvetliť alebo si vysvetlenie vymyslíme. Tieto neurónové siete môžu byť zodpovedné za celú triedu nevedomých javov, ako sú automatické činnosti, nevedomé učenie, „priming“, subliminálna percepcia, atď. Taktiež môžu vykonávať výpočty ovplyvňujúce tú neurónovú aktivitu, ktorá vedie k uvedomovaniu.

Na záver stručné zhrnutie povedaného: Vznik reflektívneho vedomia predpokladá štyri súčasné body: 1) určitú kvantitu neurónov (zjednodušene povedané, dostatočne veľký mozog), 2) ich rýchlu a premenlivú synchronizáciu v rámci jedného mozgu, 3) existenciu a funkčnosť evolučne novších mozgových oblastí, ktoré súvisia s pozornosťou a plánovaním činnosti, a 4) spoločenskú interakciu a komunikáciu. Povaha vynorenia vedomia je však stále záhadou. Je možné, že reflektívne resp. sekundárne vedomie je vyhradené iba človeku. Ale nie je vylúčená možnosť, že nejaké reflektívne vedomie majú aj ľudoopi, konkrétne šimpanzy a bonobovia (šimpanzy bonobo), prípadne iné tvory ako delfíny a slony [23]. Napríklad, ľudoopi, delfíny aj slony pri pohľade do zrkadla vedia, že sú to oni a nie iný príslušník ich druhu. Ak vedomie naozaj závisí na mozgu, nemali by sme ho upierať iným živočíchom.

Je možné, že tak ako sa na určitom stupni vývoja a organizácie hmoty objavil život, tak sa na určitom stupni vývoja a organizácie živej mozgovej hmoty objavuje vedomie. Ešte však neexistuje uspokojivá teória, ktorá by vysvetlila základnú záhadu – ako je možné, že taký fyzikálny systém, ako je mozog, s danou architektúrou a vlastnosťami, umožňuje vznik kváliám, subjektívnym pocitom a subjektívnemu prežívaniu. Stvoriť vedomie môže byť rovnako ťažké ako stvoriť život.

Poďakovanie: Tento príspevok vznikol za podpory grantovej agentúry VEGA v rámci grantovej úlohy 1/2045/05.

Literatúra

[1] Beňušková, Ľ.: Kognitívna neuroveda. In: *Kogitívne vedy,* J. Rybár, Ľ. Beňušková, V. Kvasnička (eds). Bratislava: Kalligram 2002, pp. 47-104.

[2] Beňušková, Ľ.: Kde sa jazyk stretáva s vedomím. In: *Jazyk a kognícia,* J. Rybár, V. Kvasnička, I. Farkaš (eds). Bratislava: Kalligram 2005, pp. 235-261.

[3] Chalmers, D. J.: *The Conscious Mind: In Search of a Fundamental Theory*. Oxford: Oxford University Press 1996.

[4] Crick, F., Koch, C.: Are we aware of neural activity in primary visual cortex? *Nature* **375** (1995) 121-123.

[5] Dennett, D. C.: *Consciousness Explained*. New York: Penguin Books 1991.

[6] Edelman, G. M., Tononi, G.: *Consciousness. How Matter Becomes Imagination*. London: Penguin Books 2000.

[7] Fries, P., Roelfsema, P. R., Engel, A. K., König, P., Singer, W.: Synchronization of oscillatory responses in visual cortex correlates with perception in interocular rivalry. *Proc. Natl. Acad. Sci. USA* **94** (1997) 12699-12704.

[8] Frith, U.: Mind blindness and the brain in autism. *Neuron* **32** (2001) 969-979.

[9] Kandel, E. R., Schwartz, J. H., Jessell, T. M.: *Principles of Neural Science* (3rd edition). Norwalk: Appleton & Lange 1991.

[10] Kaplan, B. J., Sadock, V. A.: *Comprehensive Textbook of Psychiatry* (*7th edition*). Philadelphia, PA: Lippincott/Williams & Wilkins 1999.

[11] Koch, C., Crick, F.: Some further ideas regarding the neuronal basis of awareness. In: *Large-Scale Neuronal Theories of the Brain*, C. Koch, J. L. Davis (eds). Cambridge, MA: MIT Press 1994.

[12] Koch, C.: Towards the neuronal substrate of visual consciousness. In: *Towards a Science of Consciousness: The First Tucson Discussions and Debates*, S. R. Hameroff, A. W. Kaszniak, A. C. Scott (eds). Cambridge, MA: MIT Press 1996.

[13] Koch, C., Hepp, K.: Quantum mechanics in the brain. *Nature* **440** (2006) 611-612.

[14] Kováč, L.: Ľudské vedomie ako produkt evolučnej eskalácie emocionálneho výberu. In: Zborník z 3. Slovensko-Českého seminára: *Kognícia, umelý život a počítačová inteligencia, CALCI 2003* (15. - 17. máj 2003, Stará Lesná), P. Sinčák, V. Kvasnička, J. Kelemen, J. Pospíchal, P. Návrat (eds), str. 21-41.

[15] Leopold, D. A., Logothetis, N. K.: Multistable phenomena: changing views in perception. *Trends in Cognitive Science* **3** (1999) 254-264.

[16] Libet, B.: Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behavioral and Brain Sciences* **8** (1985) 529-566.

[17] Libet, B.: Do we have free will? *Journal of Consciousness Studies* **6** (1999) 47-57.

[18] Llinás, R. R., Ribary U.: Perception as an oneiric-like state modulated by senses. In: *Large-Scale Neuronal Theories of the Brain*, C. Koch and J. L. Davis (eds). Cambridge, MA: MIT Press 1994.

[19] Massimini, M., Ferrarelli, F., Huber, R., Esser, S. K., Singh, H., Tononi, G.: Breakdown of cortical effective connectivity during sleep. *Science* **309** (2005) 2228-2232.

[20] Melzack, R. (1999): Phantom limb. In: *The MIT Encyclopedia of the Cognitive Sciences*, R. A. Wilson and F. Keil (eds). Cambridge, MA: MIT Press 1999.

[21] Mogilner, A., Grossman, J. A. I., Ribary, U., Joliot, M., Volkmann, J., Rapaport, D., Beasley, R. W. Somatosensory cortical plasticity in adult humans revealed by magnetoencephalography. *Proc. Natl. Acad. Sci. USA* **90** (1993) 3593-3597.
[22] Penrose, R.: *Shadows of the Mind: A Search for the Missing Science of Consciousness*. Oxford: Oxford Univ. Press 1994.
[23] Plotnik, J. M., de Waal, F. B. M., Reiss, D.: Self-recognition in an Asian elephant. *Proc. Natl. Acad. Sci. USA* **103** (2006) 17053-17057.
[24] Posner, M. I., Rothbart, M. K.: Constructing neuronal theories of mind. In: *Large-Scale Neuronal Theories of the Brain*, C. Koch and J. L. Davis (eds), Cambridge, MA: MIT Press 1994.
[25] Ribary, U., Ioannides, K., Singh, K. D., Hasson, R., Bolton, J. P. R., Lado, F., Mogilner, A., Llinas, R.: Magnetic field tomography of coherent thalamocortical 40-Hz oscillations in humans. *Proc. Natl. Acad. Sci. USA* **88** (1991) 11037-11401.
[26] Rizzolatti, G., Fadiga, L., Gallese, V., Fogassi, L.: Premotor cortex and the recognition of motor actions. *Cognitive Brain Research* **3** (1996) 131-141.
[27] Rodriguez, E., George, N., Lachaux, J.-P., Martinerie, J., Renault, B., Varela, F. J. Perception´s shadow: long-range synchronization of human brain activity. *Nature* **397** (1999) 434-436.
[28] Sahraie, A., Weiskrantz, L., Barbur, J. L., Simmons, A., Williams, S. C. R., Brammer, M. J.: Pattern of neuronal activity associated with conscious and unconscious processing of visual signals. *Proc. Natl. Acad. Sci. USA* 94 (1997) 9406-9411.
[29] Searle, J.: *Consciousness and Language*. Cambridge, MA: Cambridge Univ. Press 2002.
[30] Singer, W.: Putative function of temporal correlations in neocortical processing. In: *Large-Scale Neuronal Theories of the Brain*, C. Koch and J. L. Davis (eds), Cambridge, MA: MIT Press 1994.
[31] Singer, W.: The observer in the brain. In: *Understanding Representation in the Cognitive Sciences*. A. Riegler, M. Peschl, A. von Stein (eds), Kluwer Academic/Plenum Publishers: New York 1999.
[32] Tononi, G., Edelman, G. M.: Consciousness and complexity. *Science* **282** (1998) 1846-1851.
[33] Tye, M.: Qualia. In: *The Stanford Encyclopedia of Philosophy (Summer 2003)* E. N. Zalta (ed). http://plato.stanford.edu/archives/sum2003/entries/qualia/
[34] Žucha, I.: Psychické a kvázi-psychické: ťažkosti s pojmami. Zborník prednášok zo seminára *Kognitívne vedy IV*, Ľ. Beňušková, V. Kvasnička, J. Pospíchal (eds), str. 60-64, 2001.

Význam emocí pro umělé bytosti

Michal Bída, Rudolf Kadlec, Cyril Brom[1]

Abstrakt. Ačkoli jsou emoce neodmyslitelnou součástí kognitivních procesů lidí a pravděpodobně i vyšších zvířat, při vytváření počítačových modelů organismů jsou uplatňovány pouze zřídka. V příspěvku uvedeme některé psychologické teorie vysvětlující vznik a význam emocí a poté budeme z informatického pohledu diskutovat přínosy umělých emocí pro modely organismů. Předložené teze budeme ilustrovat na dvou původních projektech.

1 Úvod

Emoce provázejí téměř každý aspekt naší činnosti. Díky nim víme, co se nám líbí a čemu se příště raději vyhneme. Usměrňují naše myšlení, ovlivňují pozornost a jsou nám zdrojem motivace. Vnější projevy emocí, jako třeba výraz obličeje nebo držení těla, dávají našemu okolí tušit, v jaké jsme náladě a jak s námi pro tuto chvíli jednat.

Je přirozené klást si otázku, zda by se mechanismy podobné emocím nedaly použít i ku prospěchu autonomních agentů případně počítačových systémů obecně. Možný přínos emocí vidíme zejména ve dvou oblastech:

- *Zlepšení výkonu autonomních agentů působících ve vysoce proměnlivých a nepřátelských prostředích*: Autonomním agentem rozumíme počítačový systém fungující v nějakém prostředí a schopný samostatně provádět akce tak, aby splnil úkoly, ke kterým byl zkonstruován [1]. Takovýmto agentem může být robot, postava z počítačové hry nebo simulovaná myš. Jelikož jsou proměnlivá a nepřátelská prostředí podobná našemu světu, kde emoce vznikly jako jeden z mechanismů adaptace [2] (str. 20), je logické domnívat se, že by mohly pomoci i autonomním agentům.
- *Zlepšení interakce člověka s počítačem* („Human Computer Interaction“ – HCI): Za prvé může být výhodné učinit počítačové systémy obecně schopné pracovat s lidskými emocemi. Umět odhadovat emoce člověka pracujícího se systémem a následně adekvátně zareagovat by mohlo pomoci zvýšit efektivitu práce stejně jako uživatelovo pohodlí. Za druhé může být v určitých případech výhodné vytvářet určité softwarové systémy tak, aby ony samy projevy emocí imitovaly. Například v počítačových hrách by mohla tato imitace zlepšit uvěřitelnost simulovaných postav.

V prvním případě se skutečnými emocemi inspirujeme a vytváříme nové, umělé emoce. Umělé emoce se snaží postihnout základní charakteristiky a funkce skutečných emocí, ale jsou přizpůsobeny prostředí agenta a úkolům, které má vykonávat. V druhém případě se snažíme imitovat projevy lidských emocí nebo emoce podle jejich fyziologických projevů odhadovat.

[1] Katedra software a výuky informatiky, Matematicko-fyzikální fakulta Univerzity Karlovy v Praze, Malostranské nám. 25, Praha 1, 118 00, E-mail: michal.bida@seznam.cz, brom@ksvi.mff.cuni.cz.

V sekci 2 budeme nejprve hlouběji analyzovat funkce emocí v přírodě a zamyslíme se nad tím, které z nich by se daly přenést do prostředí agentů. Dotkneme se přitom vybraných teoretických modelů pracujících s emocemi ze světa psychologie i informatiky. Možné přínosy emocí budeme analyzovat ze dvou hlavních hledisek – z hlediska HCI a z hlediska konstruování autonomních agentů. V sekci 3 poté ukážeme naše dva původní modely. První z nich je zaměřený na vytváření věrohodného chování protivníka z počítačové hry. Na druhém zkoumáme možný význam emocí pro učení.

2 Emoce a stroje

Vymezení pojmu emoce bývá často ovlivněno předmětem zkoumání a občas vede k nejasnostem. Pro nás budou *emoce* (podle [3]) vyjadřovat subjektivní vztah jedince k předmětům a jevům objektivní reality i k sobě samému, který je prožíván v široké škále pocitů od výrazně kladných, přes neutrální, až k výrazně záporným pocitům.

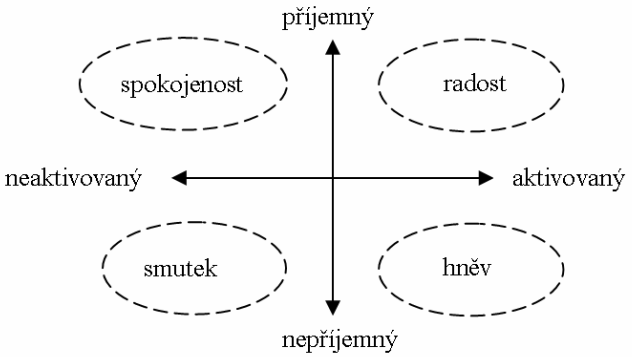
2.1 Přístupy ke zkoumání emocí

V současnosti jsou patrné dvě hlavní tendence ve studiu emocí – diskrétní přístup a dimensionální přístup.

Diskrétní přístup [4] popisuje vlastnosti jednotlivých funkčně oddělených (diskrétních) emocí. Emoce dělí na primární a sekundární. Primární chápe jako vrozené a většinou mezi ně řadí štěstí, hněv, smutek, strach, znechucení a zahanbení. Sekundární emoce pak vznikají směsicí primárních emocí a mohou být naučené, příklady jsou žárlivost, láska nebo závist.

Naproti tomu *dimensionální přístup* [5] (str. 14-15) pracuje s emocionálním prostorem definovaným pomocí různých dimenzí. Nejčastějšími dimenzemi jsou inten-zita (aktivovaný – neaktivovaný) a valence (příjemný – nepříjemný). Emocionální stav je pak charakterizován polohou v emocionálním prostoru. Diskrétní emoce můžeme chápat jako určité oblasti v tomto prostoru.

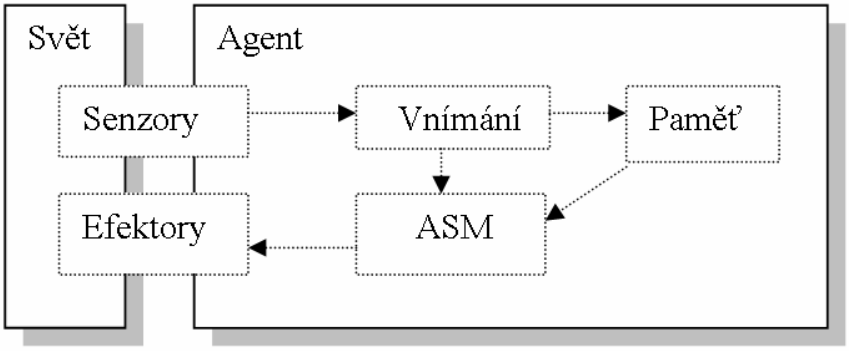
Ani jeden z výše zmíněných přístupů není universální teorií. Jsou to rozdílné pohledy na věc, které se spíše doplňují.



Obrázek 1. Vztah dimenzionálního a diskrétního přístupu.

2.2 Možný přínos emocí pro autonomní agenty

Jaký může být přínos emocí pro autonomní agenty? Pro zodpovězení této otázky nejdříve zavedeme zjednodušený a velmi obecný model agenta, podíváme se, z jakých vrstev se skládá a jak pracuje. Poté se inspirujeme funkcí emocí v živých organismech a navrhneme jejich analogické uplatnění v našem modelu.



Obrázek 2. Typická architektura agenta

Obecnou architekturu agenta a jeho pracovní cyklus ilustruje obrázek 2. Agent nejdříve zjistí stav prostředí pomocí senzorů. Proces vnímání přeloží údaje ze senzorů do reprezentace, která je vhodnější pro další rozhodování (např. nalezení určitých objektů na obrázku z kamery). Na základě těchto údajů a faktů z paměti pak mechanismus výběru akcí („Action Selection Mechanism“ – ASM) vybere nejvhodnější akci, kterou agent provede svými efektory. Učící se agent navíc vyhodnotí efekt provedených akcí a podle toho upraví paměťové záznamy nebo další parametry svého modelu.

Nyní probereme, v čem a jak by mohly emoce na jednotlivých úrovních (vnímání, ASM, paměť a učení) výše popsaného modelu agenta pomoci. Jde spíše o nástin možností, nikoli o kompletní výčet.

Usměrnění vnímání a uvažování. U lidí je usměrnění vnímání a uvažování patrné zejména na vlivu strachu a hněvu. Tyto emoce způsobují, že některé podněty z okolí nevědomě přehlížíme, zatímco jiným věnujeme veškerou naši pozornost. Tento stav bývá označován jako tzv. „tunelové vidění“ [5] (str. 89-93). Zároveň se zužuje i náš akční repertoár.

Pro agenta mohou emoce plnit funkci filtru, který například v situaci vyvolávající strach potlačí nepodstatné podněty a umožní tím věnovat více výpočetní kapacity na hledání únikových cest nebo zdrojů potenciálního nebezpečí. Jednoduchý emoční filtr, který mění důležitost objektů podle nálady, se objevuje například v práci [6].

Rychlost ASM. U lidí je emoční zhodnocení nastalé situace velmi rychlé a předchází vyšším kognitivním procesům [7]. To umožňuje téměř okamžitou reakci na základě

obecných rysů situace. Pozdější vědomé zhodnocení pak většinou naše jednání určitým způsobem koriguje.

Pro agenty může být tento mechanismus užitečný v prostředích, kde je důležitá rychlost reakce. Agent si v situaci ohrožení nemůže dovolit dopodrobna rozpracovávat plán úniku; musí zareagovat okamžitě. Pokud se bude cítit ohrožen, může okamžitě začít provádět obecný a tudíž jen přibližný plán úniku (pohybuj se ve směru od nebezpečí) a teprve potom jej začít zpřesňovat (najdi dveře a ukryj se za ně). Emoce tak mohou sloužit jako heuristika.

Vztah emocí a paměti. Experimenty potvrzují, že si s větší pravděpodobností vybavíme událost, kterou jsme si zapamatovali ve stavu odpovídajícímu našemu současnému rozpoložení [5] (str. 70-79). Tento jev bývá označován jako efekt kongruence s náladou. O jeho vysvětlení se pokusil například Bower [8]. Bower vidí paměť jako síť různých uzlů. Některé uzly odpovídají předmětům reálného světa (např. hrnek), jiné jsou svázány s emocemi (strach, štěstí atd.). Spoje mezi uzly vyjadřují, s jakou pravděpodobností si vybavíme předmět nebo emoci uzlu, pokud bude aktivní sousední uzel. Při souhlasné aktivaci dvou uzlů se spojení mezi nimi posílí. Jde vlastně o asociativní síť, ve které vystupují i emoce.

S ohledem na agenty vede existence tohoto jevu k myšlence využití emocí jako kontextu pro vyhledávání v paměti. Uzel pro hrnek může být aktivován tím, že agent právě vidí hrnek stát na stole. Pokud bude mít zároveň strach, spojení mezi uzly „hrnek“ a „strach“ se posílí. Aktivace se spojením může přelévat v obou směrech. Příště emoce strachu zvýší šanci, že si agent vzpomene na hrnek a naopak vizuální vjem hrnku u něj může vyvolat strach. V prvním případě emoce pomohou vybavit si vzpomínky získané v situaci, která byla alespoň v nějakém aspektu (zde stejném emocionálním stavu) podobná té současné. Agent si pod vlivem strachu začne vybavovat, co všechno zažil, když měl strach minule, a třeba mu to pomůže nalézt řešení jeho situace. V druhém případě vnější situace ovlivní emoce agenta a tím i jeho chování.

Učení a propojování znalostí. Pozitivní emoce pomáhají propojit nově získanou zkušenost se starou zkušeností, přispívají k nalézání souvislostí. Negativní emoce tuto vlastnost nemají, zkušenost pak často není zasazena do širšího kontextu. Tím se stane nedostupnou pro uvažování a bude zapomenuta [2] (str. 123).

Ve světě agentů se emocionální zhodnocení dá použít jako zdroj odměny při zpětnovazebním učení. To platí zejména pro prostředí, kde efekt akcí není známý. Příkladem je projekt CMatie [9].

Emoce by také mohly být použity jako heuristika, která nám řekne, kolik času se vyplatí věnovat kategorizaci nové zkušenosti.

Výše popsané přístupy předpokládaly přítomnost nějakého *emočního generátoru*, modulu, který na základě vnitřního stavu agenta a informací o stavu prostředí určí emocionální zhodnocení situace. Oprávněnost existence takového centra je podpořena neurologickými nálezy, které potvrzující důležitou roli amygdaly v emočním prožívání [7] (str. 72-75).

V informatice se objevuje ale i opačný přístup, který emoce chápe jako emergentní jevy složitého systému [10]. Například Minsky ve své koncepci mysli jako společenství

specializovaných výpočetních jednotek („cloud of resources“) chápe emoce jako vyjádření stavu těchto jednotek (rádců) [11]. Emoce podle něj odpovídají aktivaci různých rádců. Pokud o agentovi řekneme, že je rozezlený, není to podle Minského proto, že by někde v jeho modelu byl nastaven přepínač hněvu na „zapnuto“, ale proto, že je aktivní určitá podmnožina rádců a zbylí rádci jsou potlačeni.

2.3 HCI

HCI („Human-computer Interaction“) je interdisciplinární obor spojující informatiku s mnoha dalšími vědními obory, zejména psychologií, lingvistikou a sociálními vědami. Obecně řečeno obor zkoumá, jakými způsoby člověk může interagovat a jak interaguje s počítačem. Z praktického hlediska to zahrnuje především návrh, výrobu a vyhodnocování rozhraní mezi člověkem a počítačem. Tato rozhraní mohou být jak hardwarová tak softwarová. Podrobnější definici HCI lze nalézt v [12].

V poslední době se obor HCI začíná soustřeďovat mimo jiné i na *virtuální lidi* (obrázek 3), což jsou počítačové programy, které imitují lidské chování ve virtuálním prostředí, a které disponují virtuálním tělem [13]. S virtuálními lidmi též úzce souvisí problematika emocí. Intenzivní výzkum v této oblasti probíhá například v rámci evropské sítě HUMAINE [14].

Obrázek 3. Příklad virtuálních lidí z virtuálního dramatu Façade. Přejato z [15].

Využití emocí z hlediska HCI můžeme rozdělit do dvou hlavních kategorií:

Schopnost odhadovat a kategorizovat emoce. Přínos odhadování a kategorizování emocí tkví v možnosti počítačových systémů (virtuálních lidí i aplikací obecně) přizpůsobit se emočnímu rozpoložení uživatele a zvýšit tak efektivitu práce, zábavy nebo učení.

Analýza emocí se dá provádět například na základě čidel monitorujících fyziologický stav člověka nebo na základě textových promluv [20].

Schopnost imitovat emoce. Ve výukových simulacích, v počítačových hrách a v jiných aplikacích pracujících s virtuálními lidmi hraje klíčovou roli termín *uvěřitelnost*. Tento pojem má mnoho rovin. V kontextu emocí se jím rozumí míra toho, jak dobře daná postava imituje emoční jevy – tedy jestli je schopna transparentně vyjádřit svůj vnitřní emoční stav (obecně se termínem uvěřitelnost a podmínkami, které musí virtuální člověk splňovat, aby působil přesvědčivě, zabývá práce [16]).

V současné době probíhají například výzkumy zabývající se automatickým vyjadřováním emocí pomocí gestikulace a obličejového výrazu [17]. S tím souvisí otázka, jakým způsobem taková gesta nebo výrazy reprezentovat. V rámci skupiny HUMAINE probíhá projekt [19] s cílem vytvořit univerzální způsob zápisu těchto projevů (základem je jazyk XML). Emoce ovšem mohou vyjadřovat nejen virtuální lidé, ale i roboti, jak ukazuje například projekt Kismet [18].

Obě zmíněné schopnosti spolu souvisí a mohou se doplňovat. Zejména pro autonomní agenty přímo interagující s člověkem by byly přínosné obě. Například pro graficky vtělené konverzační agenty (ECA – „Embodied Conversational Agent“) a pro „chatovací“ agenty (dorozumívající se s uživatelem jen pomocí textové konzole) by bylo výhodné umět jak rozpoznávat emoce uživatele tak vyjadřovat emoce vlastní. Vyjadřování emocí zde může zahrnovat doprovodnou gestikulaci a obličejové výrazy při mluveném projevu (ECA) nebo změnu stylu a obsahu textových odpovědí (chatovací agenti).

Hlavní přínosy emocí pro obor HCI tedy vidíme zejména ve dvou oblastech. Zaprvé ve zlepšování uvěřitelnosti projevů virtuálních lidí ve výukových aplikacích, počítačových hrách a podobných programech. Zadruhé ve zlepšování schopnosti počítače rozpoznávat a odhadovat emoce člověka a na základě tohoto přizpůsobovat své chování.

3 Řešené projekty

V této sekci představíme dva původní projekty, které při modelování agentů využívají emoce. První projekt je používá ke zvýšení věrohodnosti protivníka z počítačové hry a spadá tedy pod kapitolu HCI. Druhý projekt se týká umělé inteligence; zkoumá možné provázání emocí a procesu učení.

3.1 Projekt UT Emotion Bots

Projekt emočních botů si kladl za cíl vytvořit uvěřitelný emoční model pracující v postavě v akční videohře. Akční hry typicky simulují dynamická a nedeterministická 3D prostředí (města, staré továrny ap.), ve kterých se pohybují antropomorfní postavy ovládané buď hráči nebo počítačem (viz obr. 4). Těmto postavám se říká *boti*. Cílem hry je splnit určitý úkol – dobýt základnu, získat vlajku soupeře atp. V této sekci budou popsány základní vlastnosti vyvinutého emočního modelu a nastíněny závěry prvních experimentů. Podrobnější shrnutí lze nalézt v práci [21].

Celková uvěřitelnost botů v akčních hrách je ovlivněna zejména dvěma faktory: zaprvé kvalitou grafického zpracování bota (jak dobře chodí, otáčí se, apod.), zadruhé

uvěřitelností botových akcí z hlediska jeho motivů (například když je bot zraněn, měl by se pokusit utéct).

Grafické zpracování bota závisí zejména na vybrané platformě a v našem projektu mu nebyla věnována velká pozornost (vytváření kvalitních animací je časově náročné a vyžaduje dobré programové vybavení). Projekt se místo toho soustředil na vytvoření emočního modelu, který rozšířil běžné motivy botů z akčních her o aspekt emocí, jimiž jinak postavy z her většinou nedisponují.

Výsledkem projektu jsou mimo vlastní prototyp bota s emočním modelem také poznatky o tom, na jaké aspekty emočního modelu se při vývoji soustředit. Předjímáme, že se jedná o způsob kategorizace emocí a o adaptační schopnosti modelu. Na závěr sekce budou tyto aspekty rozebrány podrobněji.

Jako testovací prostředí pro bota jsme zvolili 3D svět z akční hry Unreal Tournament (UT). Výhoda UT je v tom, že má částečně otevřený kód, díky čemuž umožňuje kontrolovat agenta externími mechanismy, a že obsahuje bohatou škálu různých prostředí.

Projekt dále využívá softwarové platformy POGamUT [22]. V rámci ní je k prostředí UT připojena reaktivně plánovací architektura POSH („Parallel-rooted, Ordered Slip-stack Hierarchy“) [23], v níž je napsaná řídící struktura bota. Vlastní emoční model je pak psán v jazyce Python [24].

Emoční model vychází z modelu A. J. Champandarda [25]. Jádrem modelu je, že všechny emoce jsou zjednodušeny na několik základních emocí dle Plutchikova kruhového modelu [4] (s drobnými úpravami). V rámci našeho modelu se pracuje se třemi základními entitami: emocemi, pocity a city:

Obrázek 4. Prostředí hry Unreal Tournament. Přejato z [26].

- *Emoce* – reprezentují déle trvající stav „mysli“ bota, pomocí nich zjišťujeme, jak se bot momentálně cítí (zda je šťastný, smutný apod.). Model obsahuje osm emocí ve čtyřech komplementárních párech – strach („fear“) a hněv („anger“), radost („joy“) a smutek („sorrow“), pýchu („pride“) a stud („shame“) a pobavenost („amusement“) a utahanost („weariness“) . Emoce se mohou nacházet v základním stavu, středně

excitovaném stavu a maximálně excitovaném stavu. V rámci komplementárního páru může být v excitovaném stavu jen jedna emoce, druhá je automaticky v základním stavu.

- *Pocity* – zprostředkovávají okamžitou reakci na stimuly z prostředí – lze je tedy chápat jako nízko-úrovňové. Na základě aktivních pocitů se poté mění emoce. Model obsahuje osm pocitů – překvapení („surprise“), očekávání („anticipation“), přitažlivost („attraction“), odpor („disgust“), bolest („pain“), potěšení („pleasure“), zmatenost („confusion“) a objevení („discovery“ – pocit při objevení nějakého nového objektu).
- *City* – reprezentují vztah bota k jiným objektům ve hře (především k ostatním hráčům a botům). City slouží jako další dodatečné hledisko při generování pocitů na základě herních událostí. Model obsahuje 4 city – přitažlivost („attraction“), odpor („disgust“), nenávist („hatred“) a lítost („pity“).

Model pracuje následovně: Přímá reakce na změny v prostředí je zprostředkována pocity. Na základě pocitů se poté generuje případná změna emocí. Vždy dva aktivované pocity způsobují s určitou pravděpodobností zvýšení jedné emoce. Po zvýšení emoce jsou příslušné dva pocity nastaveny zpět na výchozí hodnotu nula.

City jsou asociovány k určitým objektům ve hře. V přítomnosti těchto objektů, jsou na základě citů generovány další nízko-úrovňové pocity, jež mohou následně ovlivnit emoce. City jsou generovány na základě statistik, které si bot během hry ukládá. Ve statistikách figurují i některé neměnné atributy, uveďme například přitažlivost (míra toho, jak se botovi daný objekt líbí). City jako takové představují vrstvu, pomocí které se model adaptuje na prostředí.

Ilustrujme nyní fungování modelu na příkladu. Řekněme, že bot byl zasažen střelou a je zraněn, aniž by v okamžiku zásahu viděl nějakého protivníka. Z prostředí UT přijde zpráva o zranění, na kterou se zareaguje vygenerováním pocitu „bolest“. Jelikož bot nikoho neviděl, kromě pocitu bolest se vygeneruje také pocit „překvapení“. Kombinace pocitů bolesti a překvapení může situace s určitou pravděpodobností způsobit zvýšení excitace emoce „hněv“.

Čím častěji je hněv stimulován, tím pravděpodobněji může přejít do vyššího excitovaného stavu. Na druhou stranu pokud by byla stimulována komplementární emoce ke hněvu – „strach“ – došlo by ke snížení pravděpodobnosti přechodu do vyššího stavu emoce hněv a ke zvýšení pravděpodobnosti přechodu do vyššího stavu emoce strach. Nemůže tedy dojít k tomu, že by byly obě emoce ve vyšším stavu. Pokud jedna do vyššího stavu přejde, pak stimulace komplementární emoce způsobí nejprve snížení první emoce do základního stavu a až následně zvýšení excitace komplementární emoce.

Hlavním projevem změny aktuálních emocí je změna chování bota. Tato změna může být velmi jednoduše implementována díky architektuře POSH. Změna chování zahrnuje zejména změnu způsobu, jakým bot vykonává jednotlivé akce. Konkrétně se jedná například o změnu stylu mluvy bota, změnu v jeho prioritách, případně také změnu jeho základních vlastností (přesnosti a rychlosti). Například při excitované emoci strach se bot bude snažit utéct už při prvním spatření protivníka.

Bot může mít zároveň aktivní chování jen pro jednu emoci. Komplementární páry emocí byly proto ohodnoceny určitou prioritou, která se používá při výběru mezi dvěma emocemi se stejnou úrovní excitace.

Vyhodnocení a parametrizace modelu

Ukázalo se, že z hlediska uvěřitelnosti emočního modelu (jak model reaguje pomocí změn emocí na události ve virtuálním prostředí) jsou důležité dva aspekty návrhu modelu. Tyto aspekty ovlivňují uvěřitelnost emočního modelu zásadním způsobem. Jedná se zaprvé o zvolený způsob kategorizování emocí, zadruhé o adaptační schopnosti modelu. S těmito aspekty dále souvisí parametrizace modelu.

Kategorizace emocí. Způsob kategorizace emocí zahrnuje zaprvé způsob, jakým se bude konceptuálně reprezentovat vnitřní emoční stav bota, a zadruhé způsob, jakým se budou vnější události „emočně“ hodnotit. Vybraný způsob kategorizace zásadně ovlivní schopnost emočního modelu zachycovat určité emoční jevy a dále dobu strávenou parametrizací modelu. Různé psychologické teorie nabízejí různé způsoby kategorizace emocí. Záleží pak jen na emočním modelu, zda bude obsahovat jen jeden přístup ke kategorizaci nebo zda zkombinuje přístupů více.

Kategorizace emocí našeho modelu se odvíjí od klasifikace Champandarda [25]. Ukázalo se, že vytvořený emoční model je dostatečný z hlediska šíře emočních jevů, které je schopný ve hře zachycovat, ovšem na druhé straně jeho emoční zpracování vnějších událostí není příliš psychologicky plausibilní a čas strávený parametrizací modelu může být poměrně dlouhý [21]. Parametrizace je nutná z důvodu zamezení určitých negativních projevů použitého modelu: Zaprvé, při nevyvážené parametrizaci se některé emoce téměř neobjevují. Zadruhé, botův emoční stav může mít tendenci být labilní. S vyřešením těchto problémů úzce souvisí také adaptační schopnosti modelu.

Adaptační schopnosti modelu. Výzkumy ukazují, že adaptační schopnosti emočního modelu jsou klíčové pro zvýšení jeho uvěřitelnosti [29]. Adaptací se zde myslí změna emočních reakcí závislá na předchozích událostech v prostředí.

Při testování emočního modelu se ukázalo, že adaptační schopnosti modelu měly pozitivní vliv na celkovou uvěřitelnost (např. působilo věrohodně, když bot přednostně bojoval s hráči, kteří na něj útočili častěji). Adaptační schopnosti modelu zahrnují především vrstvu s city. Pro lepší výsledky by ale bylo vhodné adaptační schopnosti modelu dále rozšířit, například o emoční habituaci bota. Habituace by v kontextu emočního modelu UT Emotion bots znamenala postupné snižování emočních reakcí na často se opakující podněty.

Shrnutí

V rámci projektu UT Emotion Bots byly lépe stanoveny faktory, které ovlivňují uvěřitelnost bota v akčních videohrách. Dále byly lépe prozkoumány základní aspekty návrhu emočního modelu, na které je dobré se při vývoji soustředit. Přestože výsledky se primárně vztahují na oblast akčních her, domníváme se, že je lze aplikovat i v jiných typech aplikací pracujících s 3D světy s virtuálními lidmi, například ve výukových simulacích.

Mezi největší výhody emočního modelu UT Emotion Bots patří jednoduchá implementace a při úzkém spektru reprezentovaných emocí i únosně dlouhá parametrizace. Dále model obsahuje některé adaptační schopnosti (především vrstvy citů), které přispívají ke zvyšování uvěřitelnosti modelu. Největší nevýhodou je zejména nepříliš

psychologicky plausibilní (a někdy též obtížně parametrizovatelné) emoční zhodnocení externích událostí.

Celková uvěřitelnost UT Emotion Bots bota byla negativně ovlivněna zanedbáním grafické stránky problému. Na ni by bylo třeba se soustředit v reálných projektech.

Zajímavé a v jistých ohledech propracovanější alternativy k emočnímu modelu UT Emotion Bots nabízí OCC model [27], Rosemanův model [28] nebo model FLAME [29]. Začlenění některých přístupů z těchto modelů do našeho emočního modelu je námětem do budoucna.

3.2 Emoce a učení – myš Algernon

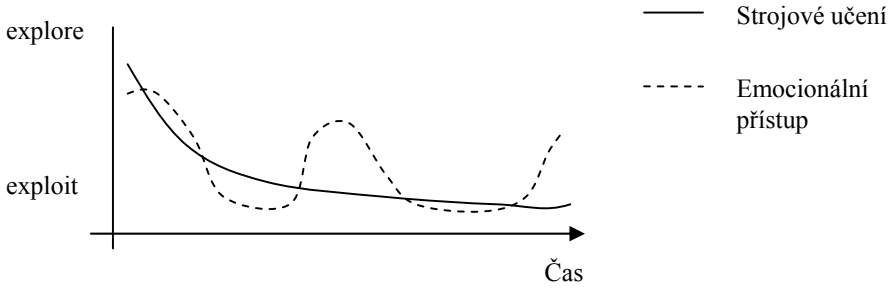
Cílem tohoto projektu bylo prozkoumat možný přínos emocí pro zpětnovazební učení („reinforcement learning“). V této sekci bude popsána základní architektura našeho modelového agenta – virtuální myši Algernon. Detaily implementace jsou popsány v [30].

Zpětnovazební učení je technika inspirovaná učením zvířat. Agent provede akci a za ni „dostane“ od prostředí odměnu nebo je potrestán. Ne vždy ale odměna přijde okamžitě, někdy je nutné k dosažení cíle provést celou sekvenci akcí. Aby takovéto sekvence mohly vznikat, je získaná odměna během učení distribuována i na předešlé akce.

Emoce u Algernon ovlivňují vyvážení mezi využíváním již nalezených strategií a průzkumem nových strategií („explore/exploit problem“). Hlavním cílem agenta není bezezbytku odhalit mechanismus fungování prostředí, agentovi stačí znát jen tolik, aby mohl úspěšně plnit zadané úkoly. Učení je pro agenta prostředkem, nikoli cílem.

S explore/exploit problémem se setkáváme i u algoritmů strojového učení. Úkolem algoritmů strojového učení je pochopit celý svět, projít všechny jeho stavy a vytvořit jeho úplný model. K tomu se často používají různé pravděpodobnostní metody, jako například simulované žíhání u neuronových sítí [31] nebo greedy algoritmus pro q-learning [32]. Jejich společným rysem je, že na začátku více prozkoumávají a ke konci už jen využívají.

Tento přístup však nemusí být pro agenty žijící v nehostinných prostředích vhodný. Fáze průzkumu může být příliš dlouhá. Agent musí brzy po začátku simulace začít uspokojovat své potřeby a nemá čas na podrobný průzkum. Idea použití emocí pro řešení explore/exploit problému je následující: agent se pokusí uspokojit svoje potřeby s tou znalostí, kterou již má, a uvidí, zdali byl skutečně úspěšný – jeho úspěšnost hodnotí (podobně jako u živých organismů) právě emoce. Pokud bude neúspěšný, prožívané emoce budou negativní a to agenta přiměje k hledání nových strategií. Rozdíl mezi navrhovaným přístupem k explore/exploit problému a přístupem běžným u strojového učení naznačuje Obrázek 5.



Obrázek 5. Závislost poměru mezi průzkumem a využíváním na čase. Strojové učení na začátku prozkoumává a postupně přechází k využívání. Emoce naproti tomu poměr adaptivně mění.

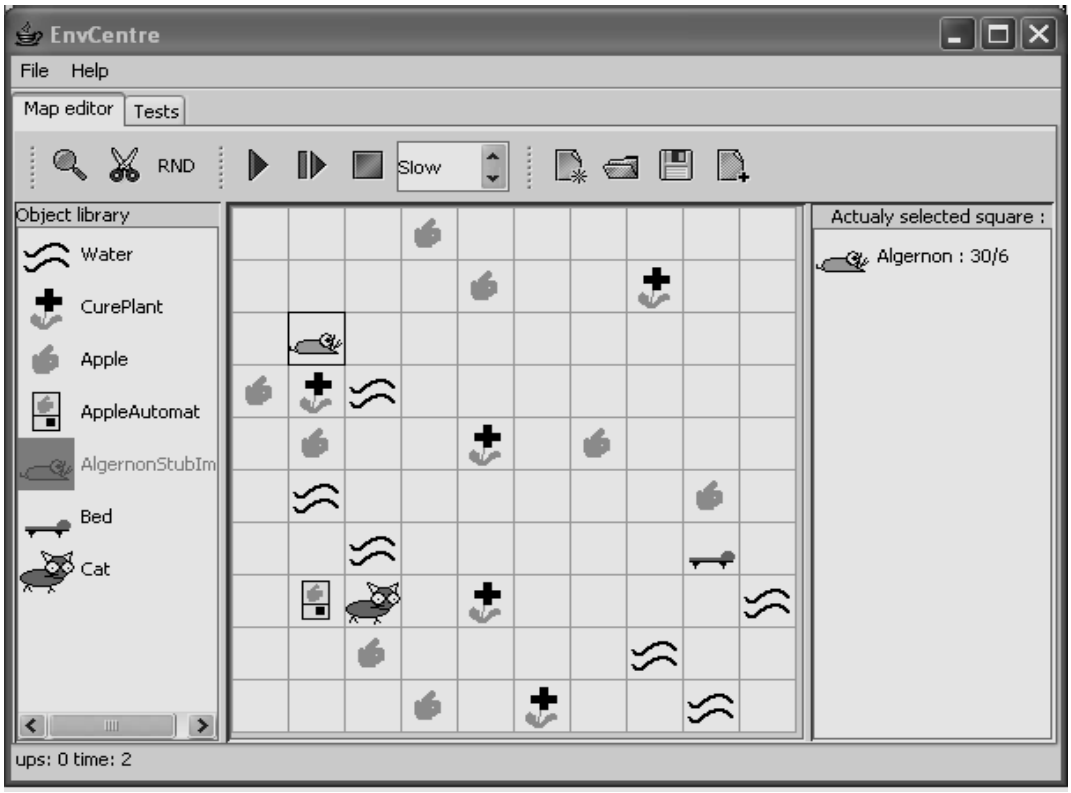
Naší tezí bylo, že adaptivní měnění poměru mezi průzkumem a využíváním, řízené emocemi, povede ke zvýšení efektivity agentova chování.

V zbylé části sekce bude nejdříve popsána neemocionální verze Algernon, její algoritmus výběru akcí a algoritmus učení a nakonec modifikovaná emocionální verze Algernon.

Algernon se snaží přežít v nehostinném světě inspirovaném prostředím Tobyho Tyrrella [33]. Naše implementace prostředí umožňuje vkládání nových, uživatelem vytvořených objektů i agentů (viz. obrázek 6). Svět a agenti fungují jako klient/server aplikace. Zvoleným programovacím jazykem byla Java.

Algernon pracuje v cyklu:

1. Vyber nejdůležitější chování.
2. Proveď akci, která je s ohledem na nejdůležitější chování nejvýhodnější.
3. Informuj paměť o odměně, již akce přinesla.



Obrázek 6. Prostředí, ve kterém je simulována Algernon

Chování odpovídají potřebám, které musí Algernon uspokojovat. Algernon se snaží získávat potravu, pít, udržovat si zdraví, odpočívat a také zkoumat své okolí. Nejdůležitější chování je vybráno na základě hodnoty aktivace – ta je funkcí interních a externích stimulů. Přirozeným interním stimulem je hodnota vnitřní proměnné, která s daným chováním souvisí. Vnitřní proměnné, s nimiž model pracuje, jsou energie, hladina vody, zdraví a únava. Například čím nižší bude energie, tím větší bude aktivace chování pro obstarávání potravy. Externí stimuly vyjadřují příležitosti, které nabízí samo prostředí.

Chování	Vnitřní proměnná
Obstarávání jídla	Energie
Obstarávání vody	Voda
Udržování zdraví	Zdraví
Odpočinek	Únava
Průzkum	bez vnitřní proměnné

Tabulka 1. Chování a jim odpovídající vnitřní proměnné

V současné verzi modelu může Algernon provádět akce: „sněz“, „napij se“, „odpočiň si“, „zmáčkni“, „jdi k“ a „jdi od“. Akce „jdi k“ a „jdi od“ mohou být provedeny na vzdálené objekty, ostatní akce jen na objekty v blízkosti myši.

Algernon si u každé akce pamatuje, jakou odměnu její provedení přinese bezprostředně a jakou odměnu může přinést v budoucnu. Okamžitou odměnou je zde chápána změna vnitřních proměnných, tedy čtveřice čísel, která odpovídá rozdílu vnitřních proměnných před a po provedení akce. Možná odměna je také čtveřice čísel; o jejich výpočtu píšeme níže.

Každý paměťový záznam má následující strukturu: [akce; typ objektu, na kterém má být akce provedena; okamžitá odměna; možná budoucí odměna].

Fungování mechanismu výběru akcí můžeme ilustrovat na následujícím příkladu. Nechť má Algernon žízeň i hlad, přičemž aktivace žízně je o něco málo větší než aktivace hladu. Vítězným chováním se tak stane pití a s ohledem na něj se bude vybírat další akce. Algernon si pamatuje, že na druhé straně mapy je tůň, a proto zvolí akci „jdi k tůni“. Nyní si představme, že Algernon cestou narazí na jablko. Jelikož se myš řídí interními i externími stimuly, může nyní dojít ke zvýšení aktivace potřeby jíst natolik, že se jedení stane nejdůležitějším chováním a pití bude odsunuto na druhé místo. To může Algernon ušetřit mnoho času a energie. Kdyby se totiž aktivace řídila jen interními stimuly, Algernon by jablko minula a pokračovala by k tůni. Teprve po uspokojení žízně, kdy by se nejdůležitějším chováním stalo obstarávání jídla, by se myš vrátila k jablku.

Aby ovšem tento jednoduchý koncept dobře fungoval, museli jsme Algernon vybavit dvěma speciálními interními stimuly: jedním je bonus za persistenci a druhým absolutní hodnota změny příslušné vnitřní proměnné (podle hesla „s jídlem roste chuť“). Konstantní bonus za persistenci si přičte pouze chování, které bylo posledním vítězem. To způsobí, že pravděpodobně vyhraje vícekrát za sebou, a proto bude moci provést delší sekvenci akcí. Tím se zvýší jeho šance na získání konečné odměny (kterou je v případě výše voda či energie získaná snědením jablka). Absolutní hodnota změny vnitřní proměnné plní podobnou funkci jako bonus za persistenci. Kdyby tyto stimuly nebyly brány v potaz, v každém kole by mohlo vyhrát jiné chování a Algernon by vyhladověla uprostřed cesty mezi jablkem a tůní.

Teď víme, jakými pravidly se řídí výběr nejdůležitějšího chování. Tento mechanismus je předem naprogramovaný a žádné učení v něm neprobíhá. Učení nastupuje na úrovni výběru akcí pro jednotlivá chování (až na chování pro průzkum, které je předem naprogramované).

Jak bylo řečeno, každému chování odpovídá jedna vnitřní proměnná (viz Tab. 1). Nejvhodnější akce se vybere podle složky odměny příslušející vybranému chování (např. pití si bude vybírat podle změny vody). Akce se seřadí podle okamžité odměny a chování se pokusí najít tu nejlepší, která se dá v současné pozici provést. Pokud se nedá provést

žádná akce, která by okamžitě přinesla kladnou odměnu (např. „napij se vody“), tak se začne vybírat podle očekávané budoucí odměny akcí. Akce se opět setřídí a vybere se ta nejlepší proveditelná (např. „jdi k vodě“).

Odměna, kterou Algernon získá provedením akce, se proporcionálně distribuuje až na pět posledně provedených akcí. U posledně provedené akce se modifikuje hodnota okamžité odměny, u starších akcí hodnota možné budoucí odměny. Nejstarší akce dostane nejmenší díl odměny, zatímco předposlední největší. To umožňuje vytvářet sekvence typu „dojdi k pasti na myši“, „zneškodni ji“ a „sněz sýr“. Tento přístup je opodstatněn myšlenkou, že předešlé akce dovedly Algernon do stavu, kde mohla provést akci přinášející skutečný užitek.

Náš algoritmus je v některých ohledech podobný známému algoritmu Q-učení [34]. Q-učení ale používá jiný mechanismus distribuce odměny a kromě samotného výsledku akcí si pamatuje i kontext, v jakém byly provedeny. Náš algoritmus je jednodušší než Q-učení, protože pracuje s menším stavovým prostorem. Kvůli tomu se Algernon nemůže naučit komplikovanější sekvence akcí, na druhou stranu učení probíhá rychleji.

Jakou roli v modelu hrají umělé emoce? Emocionální model Algernon vychází z dimenzionální teorie. Emocionální stav má jedinou dimenzi – valenci. Podmínkou pro užitečnost emocionálního modelu je, aby věrohodně hodnotil současný stav Algernon. Emocionální stav je pozitivně ovlivňován především úspěšností při plnění cílů aktivního chování, naopak negativní vliv má souběžná vysoká aktivace několika různých chování a neúspěch při jejich uspokojování – to se dá přirovnat ke stresu. Model se dále snaží postihnout některé základní charakteristiky přirozených emocí, jako například habituaci (zvykání na podněty) a saturaci (maximální hodnota jednotlivých dimenzí).

Provedené experimenty [30] naši hypotézu o užitečnosti emocí pro řešení problému explore/exploit potvrdily pouze částečně. V nepřívětivých prostředích (tedy v těch, kde se agenti dožívali nízkého průměrného věku) byla emocionální varianta Algernon lepší než neemocionální. V přívětivějších prostředích tomu bylo naopak. Tyto nejednoznačné výsledky přisuzujeme zaprvé příliš jednoduchému prostředí, ve kterém byla Algernon simulována, a zadruhé problémům s parametrizací modelu, které vedly k nestabilitě procesu učení. Nedostatek času nám bohužel nedovolil provést opakované pokusy s komplexnějším prostředím a lepší parametrizací procesu učení.

Přidáním emočního modulu jsme tak sice získali nástroj, jak ovlivňovat parametry učení a tím možná zlepšit výkon celého agenta, tento modul ovšem přináší řadu nových parametrů, jejichž manuální nastavení je nesmírně zdlouhavé. Nabízí se proto myšlenka pokusit se o automatickou parametrizaci pomocí genetických algoritmů.

4 Závěr

V tomto článku jsme přehledově shrnuli soudobé poznatky o významu emocí pro člověka a snažili se nalézt analogické využití emocí v doméně autonomních agentů a počítačových systémů obecně. Rozbor jsme prováděli ze dvou hledisek; jednak z hlediska zlepšení interakce mezi člověkem a počítačem (HCI) a jednak z hlediska lepší adaptace autonomních agentů na proměnlivá virtuální prostředí.

Předložené myšlenky jsme ilustrovali na původních projektech využívajících emoce na těchto dvou úrovních. Projekt UT Emotion Bots používá emoce jako prostředek

pro zvýšení uvěřitelnosti chování protivníků v 3D akční videohře. Druhý projekt, myš Algernon, používá emoce jako prostředek zvyšující efektivitu agentova chování.

V oboru HCI spatřujeme možný přínos emocí zaprvé ve zlepšování uvěřitelnosti projevů virtuálních lidí ve výukových aplikacích, počítačových hrách a podobných programech, a zadruhé ve zlepšování schopnosti počítače rozpoznávat a odhadovat emoce člověka a na základě toho přizpůsobovat své chování. V oboru umělé inteligence by emoce zase mohly pomoci v konstrukci agentů určených pro „reálný svět“, tedy takových agentů, kteří se budou muset potýkat se zašuměnými vstupy, nejistotou a přemírou stimulů. Je ovšem třeba podotknout, že k praktickému využití systémů s umělými emocemi ještě povede dlouhá cesta – což koneckonců ilustrují i naše projekty.

Poděkování: Chtěli bychom poděkovat všem, kteří buďto radou, nebo kritickou poznámkou přispěli k napsání toho článku, zejména Ivě Stuchlíkové, Anně Češkové a Ruth Aylett.

Tato práce byla částečně podpořena grantem GA UK 351/2006/A-INF/MFF, grantem GA UK 1053/2007/A-INF/MFF a grantem „Information Society“ pod číslem projektu: 1ET100300517.

Literatura

[1] Wooldridge, M., Jennings, N. R.: Intelligent Agents – Theories, Architectures, and Languages. In: *Volume 890 of Lecture Notes in Artificial Intelligence*, Springer-Verlag, 1995.

[2] Stuchlíková, I.: *Základy psychologie emocí*. Portál: Praha 2002.

[3] Gillerová, I.: *Slovník základních pojmů z psychologie.* Fortuna: Praha 2000.

[4] Pluchik, R.: *Emotion: A Psychoevolutionary synthesis*. Haper & Row: New York 1980.

[5] Berkowitz, L.: *Causes and Consequences of Feelings*. Cambridge University Press: Cambridge 2000.

[6] Foltýn, L.: *Realization of Intelligent Agents Architecture for Artificial Life Domain.* Master thesis, Czech Technical University in Prague, 2005, 28 – 31.

[7] Damasio, A.: *Hledání Spinozy: Radost, strast a citový mozek*. Dybbuk: Praha 2004.

[8] Bower, G. H.: Mood and memory. *American Psychologist* **36** (1981) 129-148.

[9] McCauley, L., Franklin, S.: An Architecture for Emotion. In: *Emotional and Intelligent: The Tangled Knot of Cognition*, Menlo Park, CA: AAAI Press, 1998, 122 – 127.

[10] Coutinho, E., Miranda, E. R., Cangelosi, A.: Towards a Model for Embodied Emotions. In: *Proceedings of the Workshop on Affective Computing: Towards Affective Intelligent Systems*, Covilhã, 2005.

[11] Minsky, M.: *The Emotion Machine*. New York: Simon & Schuster 2006, draft [25.1.2007], web.media.mit.edu/~minsky/

[12] ACM SIGCHI: Curricula for Human-Computer Interaction, 1992, [25.1.2007], www.sigchi.org/cdg/

[13] Brom, C., Šisler, V., Hoffmann, M.: Virtuální lidé. In: (Mařík V., Štěpánková O., Lažanský J., eds): *Umělá inteligence 5*. Academia: Praha (v tisku)

[14] HUMAINE: Human-Machine Interaction Network on Emotion. Network of Excellence in the EU's Sixth Framework Programme, in the IST (Information Society Technologies) Thematic Priority IST-2002-2.3.1.6 Multimodal Interfaces (2004-2008), www.emotion-research.net
[15] Mateas, M., Stern, A.: Façade: An Experiment in Building a Fully-Realized Interactive Drama. In.: *Game Developers Conference*, 2003.
[16] Loyall, A. B.: *Believable Agents.* Ph.D. thesis, Tech report CMU-CS-97-123, Carnegie Mellon University, May 1997.
[17] Hartmann, B., Mancini, M., Pelachaud, C.: Towards affective agent action: Modelling expressive ECA gestures. In: *Workshop on Affective Interactions at IUI 2005*, San Diego, 2005.
[18] Kismet, [23.1.2007], www.ai.mit.edu/projects/humanoid-robotics-group/kismet/
[19] Kopp, S., Krenn, B., Marsella, S., Marshall, A., Pelachaud C., Pirker, H., Thorisson, K., Vilhjalmsson, H.: Towards a Common Framework for Multimodal Generation in ECAs: The Behavior Markup Language. In: J. Gratch et al. (eds.): *Intelligent Virtual Agents 2006*, LNAI 4133, Berlin: Springer-Verlag, 2006, 205-217.
[20] Zhang, L., Barnded, J. A., Hendlez, R. J., Wallington, A. M.: Affect Detection from Open-Ended Improvisational Text. In: *Proceedings of AISB 2006*, University of Bristol, III, 78 – 85, 2006.
[21] Bída, M.: *Emoční boti v prostředí hry Unreal Tournament.* Bakalářská práce, Matematicko-fyzikální fakulta, Univerzita Karlova v Praze, 2006.
[22] Brom, C., Gemrot, J., Bída, M., Burkert, O., Partington S. J., and Bryson, J. J.: POSH Tools for Game Agent Development by Students and Non-Programmers. In: *Proceedings of CGAMES 06*, Dublin, Ireland, 2006, 126 – 135
[23] Bryson, J. J.: POSH (Parallel-rooted, Ordered Slip-stack Hierarchical) Reactive Plans, 2005, [25.1.2007], www.cs.bath.ac.uk/~jjb/web/posh.html
[24] Python, [25.1.2007], www.python.org
[25] Champandard, A. J.: *AI Game Development: Synthetic Creatures with Learning and Reactive Behaviors*. New Riders, 2003.
[26] Epic MegaGames: Unreal Tournament, 1999, [9.4.2006], www.unrealtournament.com
[27] Ortony, A., Clore, G. L., Collins, A.: *The Cognitive Structure of Emotions*. Cambridge University Press: Cambridge 1988.
[28] Roseman, I. J., Antoniou, A. A., Jose, P. E.: Appraisal Determinants of Emotions: Constructing a More Accurate and Comprehensive Theory. *Cognition and emotion* **10** (1996) 241-277.
[29] El-Nasr, M. S., Yen J., Ioerger, T.: FLAME - A Fuzzy Logic Adaptive Model of Emotions. In: *Automous Agents and Multi-agent Systems* **3** (2000) 219-257.
[30] Kadlec, R.: *Adaptivní agenti a emoce,* Bakalářská práce, Matematicko-fyzikální fakulta, Univerzita Karlova v Praze, 2006.
[31] Wikipedia: Simulated annealing - Wikipedia, the free encyclopedia [online]. 2006 [cit. 2006-05-19]. en.wikipedia.org/wiki/Simulated_annealing
[32] Richard, S.: Generalization in reinforcement learning: Successful examples using sparse coarse coding. *Advances in Neural Information Processing Systems* **8** (1996) 1038-1044.

[33] Tyrrell, T.: *Computational Mechanisms for Action Selection*. PhD thesis, University of Edinburgh, 1993.
[34] Christopher, W.: *Learning from delayed rewards*. PhD. thesis, University of Cambridge, Psychology Department, 1989.

Samoorganizace a kognice

Jan Burian[1]

Abstrakt. Článek zkoumá možnosti metodického použití pojmového aparátu teorie samoorganizace (self-organization) a systémové komplexity (complexity) k popisu kognitivních procesů. Hlavním tématem článku je rozbor různých vhodných formálních definic systémové komplexity, která je klíčovou vlastností samoorganizace, a jejich vztah ke kognici. Konkrétně rozebíráme samoorganizaci u celulárních automatů, neurální komplexitu (neural complexity) a struktury podléhající mocninným zákonům (power law), které indikují tzv. bezškálovost.

1 Samoorganizace a kognice

Samoorganizace (self-organization) je proces probíhající v otevřených autonomních systémech, při kterém se zvyšuje nebo udržuje komplexita (complexity) vnitřní organizace systému.[2]

Organizací systému rozumíme strukturu vazeb mezi částmi systému. Jedná se o prostorové uspořádání, dynamiku interakcí a vzájemný vztah mezi nimi. Autonomnost znamená, že systém není řízen z vnějšího prostředí. Otevřenost systému znamená, že systém interaguje se svým prostředím. Samoorganizující se systém tedy interaguje se svým prostředím, ale jeho stav není prostředím jednoznačně determinován.

Klíčový je pojem systémové komplexity, kterému se budeme později věnovat podrobně. Prozatím se spokojíme s vágní charakteristikou, že komplexní systém je složen z velkého množství vzájemně zpětnovazebně interagujících částí[3], a že je nesnadné popsat chování systému jako celku. Můžeme pouze více či méně popsat chování jednotlivých částí systému a jejich vzájemné interakce. Teprve na základě skládání těchto interakcí vzniká chování systému jako celku[4]. Chování celku systému je často velice obtížné, popřípadě nemožné dopředu přesně analyticky předpovídat. Je však možné celý systém s jistou mírou přesnosti kvantitativně modelovat. Je také možné zkoumat obecná pravidla, která řídí některé charakteristiky celkového chování systému.

Jako samoorganizující se systémy jsou chápány fyzikální a chemické struktury (např. perkolace, oscilující chemické reakce, autokatalytické cykly, polymerace), biologické procesy (např. skládání organických makromolekul, homeostáze, morfogeneze v rámci jedince i druhu, fungování societ organismů) i společenské procesy (např. vývoj a struktura jazyka, davové chování, fungování trhu, fungování velkých organizací, internet). Kog-

---

[1] Vysoká škola ekonomická, Fakulta informatiky a statistiky, Nám. W. Churchilla 4, Praha 3, 130 00, E-mail: burianj@vse.cz

[2] Současná terminologie je nejasná. Samoorganizace, tak jak ji zde chápeme je jedním ze způsobů chování komplexních systémů (complex systems). Komplexní systémy, jsou obecně charakterizovány složitou vnitřní strukturou a interakcí částí, přičemž ale jejich vývoj nemusí nutně směřovat k udržování komplexity. V tomto smyslu je samoorganizace spíše vlastností adaptivních komplexních systémů. V tomto článku pro jednoduchost používáme termín samoorganizující se systém.

[3] Samotná zpětnovazebnost však nemusí vést nutně k samoorganizaci. Pozitivní zpětná vazba může vést systém k neomezené divergenci, negativní zpětná vazba může naopak procesy v systému zcela utlumit.

[4] V tomto smyslu je možno samoorganizaci chápat jako pokus o nové uchopení a zpřesnění intuicí stojících za pojmem emergence.

nitivního agenta můžeme metodicky chápat[5] jako samoorganizující se systém. Takovým agentem může být jednobuněčný i mnohobuněčný organismus, část organismu (např orgán či skupina buněk), societa, druh[6], nebo i simulovaný agent v simulovaném prostředí.

Kdyby agent reagoval na všechny změny v prostředí, byl by zahlcen množstvím podnětů, z nichž většina má povahu náhodného šumu. Kdyby naopak agent nedokázal reagovat na podněty, nebo byly jeho reakce stále stejné na jeden podnět bez ohledu na jiné podněty (kontext podnětu), nedokázal by efektivně reagovat na změny prostředí. Reakce samoorganizujícího se kognitivního agenta se musí pohybovat někde mezi těmito extrémy. A jak uvidíme, tento typ chování souvisí s komplexitou.

Budeme předpokládat, že samoorganizující se agent (tj. udržující či zvyšující komplexitu) je autonomní, situovaný (situated) a vtělený (embodied). Autonomii jsme definovali výše. Situovanost souvisí s otevřeností systému a znamená, že agent je schopen učit se orientovat v prostředí, které může aktivně poznávat. Konečně vtělenost znamená, že agent má tělo, které je součástí prostředí, umožňuje poznávání prostředí a je schopné prostředí modifikovat (včetně vlastního těla). Vnitřní stavy agenta jsou stavy jeho těla.

Kognici takového agenta budeme chápat široce jako aktivní schopnost agenta rozpoznávat změny prostředí významné pro jeho přežití, reagovat na ně, adaptovat se na ně a ovlivňovat prostředí. Toto pojetí kognice se v základních rysech vyskytuje u celé řady současných teoretických biologů, přestože se mohou lišit jejich epistemologická východiska [13,15,10,11].

Jelikož je agent autonomní, nemusí reagovat na každou změnu stavu prostředí, může selektovat jen některé stavy, které jsou pro něj dle nějakých kritérií důležité. Pokud má agent zůstat autonomní, musí udržet vlastní integritu, tedy udržet složitost vnitřní organizace a schopnost samoorganizace. Kritériem pro rozlišení důležitých stavů prostředí je to, zda jsou důležité pro další existenci kognitivního agenta jako samoorganizujícího se systému. U biologických agentů, a s jistou míru nadsázky i u nebiologických, můžeme říci, že samoorganizující se kognitivní agent rozpoznává ty stavy prostředí, které jsou důležité pro jeho přežití.[7] Jelikož tělo agenta je součástí prostředí, agent by měl být schopen rozpoznávat i stavy vlastního těla důležité pro přežití.

Agent samozřejmě může (ne nutně) rozpoznávat stavy prostředí a vykonávat akce, které jsou neutrální vzhledem k jeho přežití.

Jak vnitřní organizace agenta tak jeho prostředí se může měnit, mohou se měnit i kritéria pro rozlišování stavů nutných pro přežití. Adaptace je schopnost agenta vytvářet varianty různých způsobů rozpoznávaní stavů prostředí a efektivního výběru mezi těmito způsoby. Právě výše zmíněná schopnost rozpoznávat neutrální změny prostředí podporuje adaptaci, neboť stavy, které byly dříve neutrální se mohou stát pro přežití výhodné.

Adaptace v biologických systémech existuje jak na úrovni druhu, tak na úrovni jednotlivce. Na úrovni druhu jde o vytváření neutrálních změn v genotypu, které se při

---

[5] Z metodologického hlediska je důležité neztotožňovat žádnou empiricky poznávanou část skutečnosti s jejím popisem. Tedy pojem „kognitivní agent“, „samoorganizace“ a další pojmy jsou jen pomůcky, které slouží k popisu skutečnosti, k jejímu hlubšímu pochopení, nejsou s ní však totožné.

[6] Druh je možné chápat jako samoorganizujícího se kognitivního agenta na rozsáhlých časových a prostorových škálách

[7] Co je důležité pro přežití autonomního agenta může být určeno uměle zvenčí systému. To je příklad kritérií pro šlechtění hospodářských zvířat, nebo nastavení fitness funkce při umělé evoluci. Vlastnosti agenta jsou takto sice omezeny, ale nikoliv jednoznačně determinovány, agent zůstává nadále autonomní.

pozdějších změnách prostředí mohou ukázat jako výhodné. Na úrovni jedince jde o různé typy učení pomocí pokusu a omylu. Podobné typy adaptace nacházíme i v sociálních a umělých systémech.[8]

2 Modelování komplexity

Existují systémy, které sice vykazují vlastnosti samoorganizace, avšak jejich schopnost rozlišovat stavy prostředí důležité pro její udržení je nízká. Jako příklad uveďme růst pevných krystalů nebo výše zmíněné oscilující chemické reakce. Tyto systémy buď na změny prostředí téměř nereagují díky svému velmi pravidelnému a pevnému vnitřnímu uspořádání (krystaly). Nebo jsou naopak na změny prostředí (např. množství dodávané energie) vysoce citlivé, neboť jejich vnitřní uspořádání vykazuje vysokou míru neuspořádanosti (oscilující reakce).

Do jisté míry je možné rozdíl mezi těmito dvěma krajními příklady ilustrovat vhodnými matematickými modely samoorganizace[9]. Jedny z nejpopulárnějších, nejjednodušších a zároveň nejnázornějších jsou celulární automaty. Jednorozměrný celulární automat je systém tvořený řetězcem buněk, které nabývají v hodnot (například {0;1}), přičemž stav buňky v následujícím kroku je určen stavem k jejích sousedů v kroku minulém. Počet možných pravidel, která určují jak se bude měnit stav buněk, prudce roste s v i k. [10]

Různá pravidla mají různý vliv na chování systému. Pokud například první řetězec buněk bude náhodně vygenerovaný a nové řetězce skládáme pod sebe, vzniknou obrazce znázorňující strukturu postupné (samo)organizace náhodně vygenerovaného řetězce. Wolfram [16] popisuje 4. třídy chování jednorozměrných pravidel.

První dvě třídy chování vedou k jednoduchému uniformnímu nebo repetitivnímu chování. Pokud bychom počáteční náhodně vygenerovaný řetězec chápali jako fluktuaci vyvolanou vnějším prostředím, pak bychom tyto dvě třídy chování mohli připsat vysoce uspořádaným systémům, nepříliš citlivým na vnější podněty (jako jsou pevné krystaly), kdy vnější podnět (do nějakého kritického stupně intenzity) je buď zcela utlumen nebo

---

[8] Adaptace pomocí pokusu a omylu samozřejmě vytváří i případy, které nejsou aktuálně ani výhodné ani neutrální, ale naopak způsobí zánik agenta. Těmto případů se nelze vyhnout. Samoorganizující se agent se chová tak, že takové případy jsou ze statistického hlediska akceptovatelnou daní za schopnost adaptace.

[9] Omezil jsem se zde na diskrétní modely samoorganizace. Těch je celá řada, v podstatě tvoří (s více či méně jasnými hranicemi) podmnožinu tzv. konekcionistických systémů (například rekurentní neuronové sítě s Hebbovskou nebo jinou formou učení bez učitele, samoorganizující se (Kohonenovy) mapy atributů a různé systémy založené na umělé evoluci). Tyto modely jsou však relativně složité (mění se nejen stavy systému, ale i pravidla určující jejich vývoj) a proces samoorganizace je u nich méně názorný než u celulárních automatů.

Praktické aplikace konekcionistických systémů bývají omezeny na modelování velice specifických kognitivních funkcí. Pro složitější funkce (resp. složitější množiny vstupů) je velmi obtížné navrhnout takovou architekturu systému a takové parametry jeho vývoje, které by zaručily, že bude konvergovat k efektivnímu chování. Pomoci zde mohou evoluční metody, které umožňují generování širokého spektra různých variant systému a tak zajišťují jeho adaptaci. Přesto je většina umělých konekcionistických systémů velmi vzdálená od schopností reálných samoorganizujících se kognitivních agentů. Zejména jde o schopnost nejen adaptovat se na změny prostředí, ale skutečně být součástí prostředí prostřednictvím těla, které může aktivně prostředí ovlivňovat. Absence této schopnosti znemožňuje umělým systémům efektivně modelovat složitější kognitivní funkce.

[10] Konkrétně pokud označíme v počet stavů, které mohou buňky nabývat, a k počet buněk, které určují nový stav buňky, pak existuje $v^{v^{\wedge k}}$ možných pravidel.

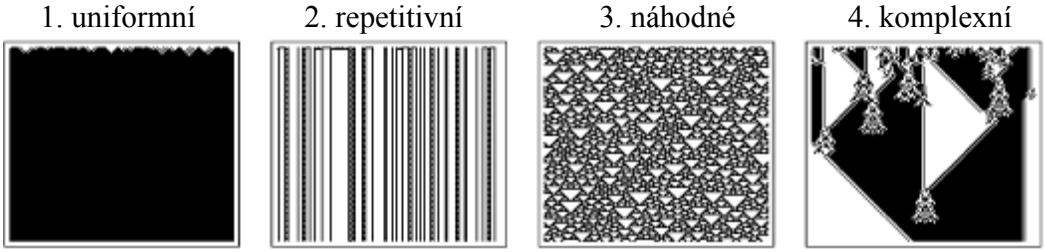
vede k jednoduché vibraci. Jako celek mohou tyto systémy v průměru nabývat jen málo odlišných stavů.

Zcela naopak je tomu ve třetím případě, kdy systém nepřestává fluktuovat, míra neuspořádanosti systému se nemění.[11] Tyto systémy jsou velmi citlivé na počáteční podmínky. Jako celek mohou tyto systémy nabývat velmi mnoha stavů. Chování této třídy nazýváme náhodné (random). [12]

Čtvrtá třída chování je někde na rozmezí mezi prvními dvěma třídami a třídou třetí. Počáteční náhodná fluktuace je relativně rychle utlumena, ale v některých ostře ohraničených lokálních strukturách přetrvává, může se systémem šířit a interagovat s jinými podobnými lokalizovanými strukturami (někdy se tyto struktury nazývají glider – kluzák). Chování této třídy velmi jednoduše modeluje chování samoorganizujícího se kognitivního agenta. Systém selektuje jen jisté stavy vnějšího prostředí (jen jistá uspořádání počáteční fluktuace), tyto stavy vedou ke změnám stavů systému (lokalizované struktury), tyto struktury mezi sebou interagují a tím vytvářejí varianty možných kontextů ve kterých se mohou objevit nové vstupy z prostředí, což je základem adaptace. Poznamenejme ještě, že interakce lokalizovaných struktur slouží k výměně informací mezi částmi systému.

Chování této třídy nazýváme komplexní. Zde nastávají jisté terminologické problémy. Systémovou komplexitu je nutno odlišit od jiných typů komplexity neboť pojem komplexity je používán v různých teoriích a označuje poněkud odlišné jevy.

Konkrétně se nebudeme zabývat komplexitou v teorii výpočetní složitosti, která zkoumá časové a paměťové nároky na řešení nějakého problému.



Obrázek 1. Čtyři třídy chování jednorozměrných celulárních automatů [16].

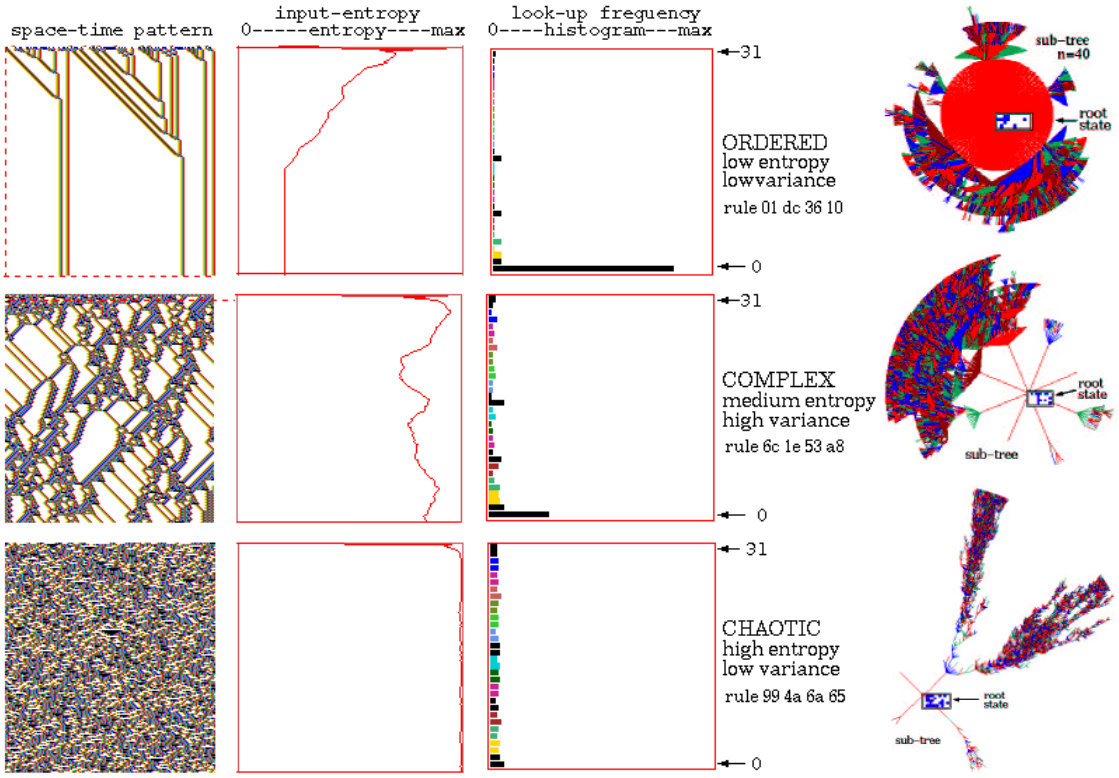
Teorie algoritmické informace zase popisuje různé míry pro měření komplexity bitových řetězců. Například Kolmogorovova komplexita udává komplexitu řetězce jako délku popisu algoritmu, který dokáže tento řetězec vygenerovat. Řetězce pro které existuje krátký algoritmus mají mnoho pravidelností a mají nízkou Kolmogorovovu komplexitu. Naopak existují řetězce, pro které nelze určit, zda popis algoritmu, který je popisuje, je kratší než délka řetězce. Takové řetězce mají vysokou Kolmogorovovu komplexitu. Tyto řetězce jsou však většinou zcela náhodné (resp. neexistuje způsob jak v nich odhalit pravidelnosti), a jsou podobné stavům systému ze 3. Wolframovy třídy chování celulár-

[11] Některá pravidla ze 3. třídy vedou k náhodným fluktuacím i ze zcela uspořádaných počátečních stavů (zvyšují neuspořádanost systému), jiná pouze udržují neuspořádanost počátečního stavu.

[12] Autoři zabývající se chováním samoorganizujících se systémů většinou rozlišují náhodné a chaotické děje. Teorie deterministického chaosu zkoumá chaotické děje, jejichž deterministický vývoj je určen soustavou nelineárních diferenciálních rovnic, které jsou citlivé na počáteční podmínky. Pravidla určující vývoj samoorganizujících se systémů však mohou být mnohem obecnější.

ních automatů. Proto je toto pojetí pro nás nevhodné. Jak upozorňuje například Gell-Mann [8], komplexita Shakespearových sonetů měřená touto mírou by byla menší než náhodné blábolení, které by vyťukávala do klávesnice psacího stroje opice.

Komplexita kognitivních procesů nemá měřit délku popisu celého systému, ale jen délku popisu pravidelností v systému. Proto by měla být komplexita nízká jak pro vysoce uspořádané systémy (ve kterých je jen málo typů pravidelností) tak pro náhodné systémy (ve kterých žádné pravidelnosti nejsou). Zde vidíme souvislost se schopností kognitivního agenta selektovat z prostředí pouze stavy důležité pro udržení samoorganizace, a to jsou právě ty stavy, které se více či méně pravidelně opakují, neboť na ně se může adaptovat.



Obrázek 2. Kolísání entropie pro uspořádané, komplexní a chaotické systémy [17]. V prvním sloupci je zachycen vývoj jednorozměrných celulárních automatů[13] s uspořádaným, komplexním a náhodným (zde nepřesně chaotickým) chováním. Ve druhém sloupci jsou zachyceny příslušné hodnoty entropie, pro komplexní chování entropie kolísá a má vysoký rozptyl. Ve třetím sloupci je zachycena distribuce možných stavů částí (pětice buněk) systému.[14] V posledním sloupci jsou ukázky struktury stavového prostoru v okolí atraktoru systému.

[13] Jedná se o celulární automaty u nichž v=2 a k=5. Buňky na obrázku jsou obarvené podle stavu pětice jejich sousedů v předchozím kroku.

[14] V tomto případě byla entropie počítána na základě frekvencí stavů ve kterých se nacházejí pětice po sobě následujících buněk. Jelikož v=2 a k=5, existuje 2^5 tj. 32 možných stavů. Tyto frekvence byly navíc ještě zprůměrovány pro deset po sobě následujících kroků.

U uspořádaných systémů je pravděpodobnostní distribuce stavů pětic výrazně asymetrická (nejnižší entropie), u komplexních je vyrovnanější a u náhodných jsou pravděpodobnosti jednotlivých stavů téměř totožné (nejvyšší entropie).

Hrubé charakteristiky systémové komplexity podává Wuensche [17], který pomocí speciálního software DDLab (Discrete Dynamics Laboratory) zevrubně zkoumal dynamiku velkého množství celulárních automatů a jejich zobecnění tzv. náhodných booleovských sítí (Random Boolean Network)[15].

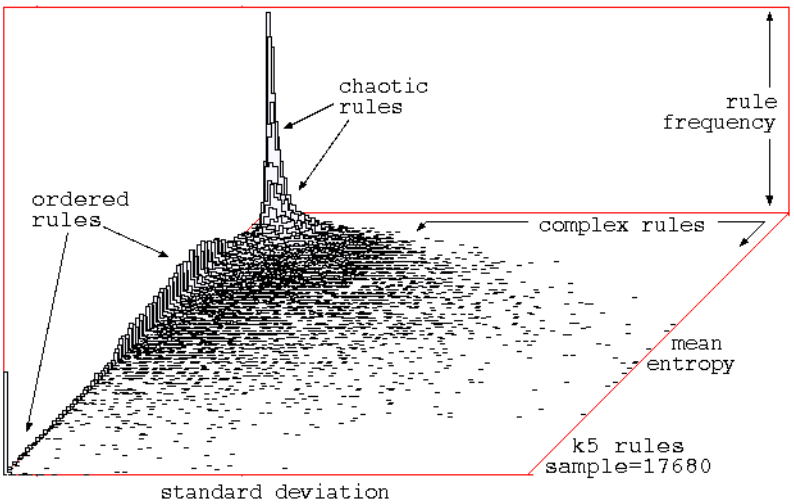
Wuensche zkoumal vývoj entropie systému v čase a empiricky ukázal, že v případě uniformních a repetitivních systémů je entropie nízká, v případě náhodných systémů je vysoká, ale pro všechny tyto třídy chování platí, že entropie zůstává při vývoji systému stabilní. Naopak komplexní chování se vyznačovalo kolísající mírou entropie.

Entropie je míra neuspořádanosti systému a znamená vlastně kolik jednotek, které mohou nabývat *r* stavů[16] průměrně potřebujeme abychom vyjádřili jeden stav systému.

$$H(X) = - \Sigma_{i=1..n}\, p_i \log_r p_i$$

Zde $H(X)$ je entropie systému (reprezentovaného pravděpodobnostní veličinou X), n je počet stavů systému a p_i pravděpodobnost i-tého stavu. Entropie roste s počtem možných stavů systému a klesá s rozptylem jejich pravděpodobností. Tedy čím více stejně pravděpodobných stavů může systém nabývat tím větší bude jeho entropie.

Kolísání entropie je způsobeno interakcemi lokalizovaných struktur. Entropie systému je přímo úměrná počtu lokalizovaných struktur. Při některých interakcích se počet lokalizovaných struktur zvyšuje, při jiných snižuje a proto entropie kolísá.



Obrázek 3. Klasifikace pravidel vývoje celulárních automatů [17]. Tento graf analyzuje 17690 náhodně vybraných pravidel z celkem 2^{32} možných pravidel pro vývoj celulárního automatu s v=2 a k=5. Výška sloupce značí frekvenci pravidel. Pravidla vedoucí ke komplexnímu chování jsou v pravé části grafu, mají vysokou směrodatnou odchylku entropie (která je mírou kolísání). Vidíme, že tato pravidla tvoří jen zlomek všech pravidel.

Komplexní chování se odráží i ve struktuře stavového prostoru. V posledním sloupci obrázku 2. je znázorněna struktura stavového prostoru v okolí jednoho z atraktorů

[15] Pomocí náhodných booleovských sítí je například možné modelovat libovolné neuronové sítě s diskrétními stavy.

[16] V informatice bývá *r* většinou rovno 2, entropie se počítá v dvouhodnotových jednotkách - bitech, obecně tomu ale být nemusí.

systémů (tzv. bazény přitažlivosti – atractor basins). Úsečky směrem do středu znázorňují přechody mezi stavy. Pokud na systém nepůsobí vnější vlivy, konverguje do stavu, který už se dále nevyvíjí (bodový atraktor) nebo do uzavřeného cyklu stavů (cyklický atraktor)[17]. Vidíme, že uspořádané systémy konvergují velmi rychle, mají vysokou frekvenci stavů s mnoha předchůdci (vysoký tzv. in-degree) a zároveň vysokou frekvenci tzv. rajských (Garden of Eden) stavů, které nemají vůbec žádného předchůdce. Náhodné systémy konvergují pomalu, mají vysokou frekvenci stavů s málo předchůdci (nízký in-degree) a málo rajských stavů. Komplexní systémy konvergují středně rychle, a závislost frekvence stavů na počtu jeho předchůdců je dána tzv. mocninným zákonem (power law). Mocninné zákony jsou další důležitou charakteristikou komplexity, podrobněji je rozebereme v příslušné kapitole.

Pravidel, které vedou ke komplexnímu chování, je jen malý zlomek ze všech možných pravidel. Například Wolfram [16] ukazuje že z 256 možných pravidel pro vývoj celulárního automaty s v=2 a k=3, pouze dvě pravidla vedou ke komplexnímu chování, které se vyznačuje interakcí lokalizovaných struktur. Již zmíněný program Ddlab umožňuje analyzovat typy pravidel v přehledném grafu.

3 Neurální komplexita

3.1 Specializace a integrace

Mezi neurovědci probíhala a stále probíhá diskuze, zda zpracování podnětů v mozku je spíše záležitostí jasně lokalizovaných a specializovaných center, nebo zda se jedná o holistický proces v němž jsou integrovány relativně rozsáhlé oblasti mozku.

Na začátku 90. let se Tononi pokusil ukázat, že v této debatě mají pravdu obě strany, a že pro efektivní zpracování podnětů musí být vzájemně na větší vzdálenost dobře integrovány rozsáhlé oblasti, které jsou ale složené z lokálně silně propojených a specializovaných center. Tyto výsledky jsou shrnuté a rozvinuté v [6].

Tononi navrhl míru komplexity založenou na vzájemné informaci (mutual information) mezi částmi systému. V případě mozkové tkáně sledujeme vzájemnou závislost vzorců neuronové aktivity. Konkrétně jde o to, do jaké míry ovlivní změna jedné části systému zbytek systému. Vzájemná informace mezi částmi systému je definována:

$$MI(X_j^k, X\text{-}X_j^k) = H(X_j^k) + H(X\text{-}X_j^k) - H(X)$$

Zde $H(X_j^k)$ je entropie j-té části systému o velikosti k, $H(X\text{-}X_j^k)$ je entropie zbytku systému a $H(X)$ je entropie celého systému. Součet $H(X_j^k)+H(X\text{-}X_j^k)$ bude vždy vyšší nebo roven $H(X)$, vzájemná informace bude tedy vždy vyšší nebo rovna nula. Entropie části systému nemůže být vyšší než entropie celého systému a proto vzájemná informace může být maximálně $H(X)$ (a to v případě že $H(X_j^k) = H(X\text{-}X_j^k) = H(X)$).

U vysoce uspořádaného systému se entropie jednotlivých částí blíží entropii systému, a tedy i vzájemná informace se blíží entropii celého systému. Jelikož je však počet stavů systému nízký, je nízká i jeho entropie, a tedy i vzájemná informace mezi částmi. Jinými slovy změna části neznamená větší změnu zbytku systému.

U náhodného systému je naproti tomu počet stavů vysoký, proto je vysoká i

[17] Jedná se o konečné diskrétní systémy, proto se zde nevyskytují tzv. podivné atraktory známé z teorie chaosu. Některé cyklické atraktory však mohou být extrémně dlouhé.

entropie celého systému. Avšak součet entropií částí se blíží k entropii celého systému a tedy vzájemná informace se blíží k nule. Změna části opět nemění zbytek systému.

Komplexní systémy mají relativně vysoký počet stavů, tedy i vysokou entropii, zároveň součet entropií částí významně přesahuje entropii celku a vzájemná informace je tedy vysoká. Změna části tedy znamená i změnu zbytku systému.

Tononi zavádí míru nazvanou neurální komplexita (neural complexity),

$$C_N(X) = \Sigma_{k=1..n/2} <MI(X_j^k, X- X_j^k)>$$

která je sumou průměrné vzájemné informace mezi částmi systému o velikosti k a zbytkem systému. Průměr je označen pomocí špičatých závorek <.>.

Tononi argumentuje, že z anatomického hlediska mají největší komplexitu vzorce neuronové aktivity v oblastech, které vykazují jak lokalizovanou specializaci, která přispívá ke zvýšení možných stavů systému, tak dobrou celkovou integraci, která umožňuje, aby změna části přispívala ke změně zbytku systému. Tato integrace navíc zajišťuje i zpětnou vazbu, takže centrum, které vyvolalo změnu u jiných center, může být po nějaké době těmito změnami samo ovlivněno.

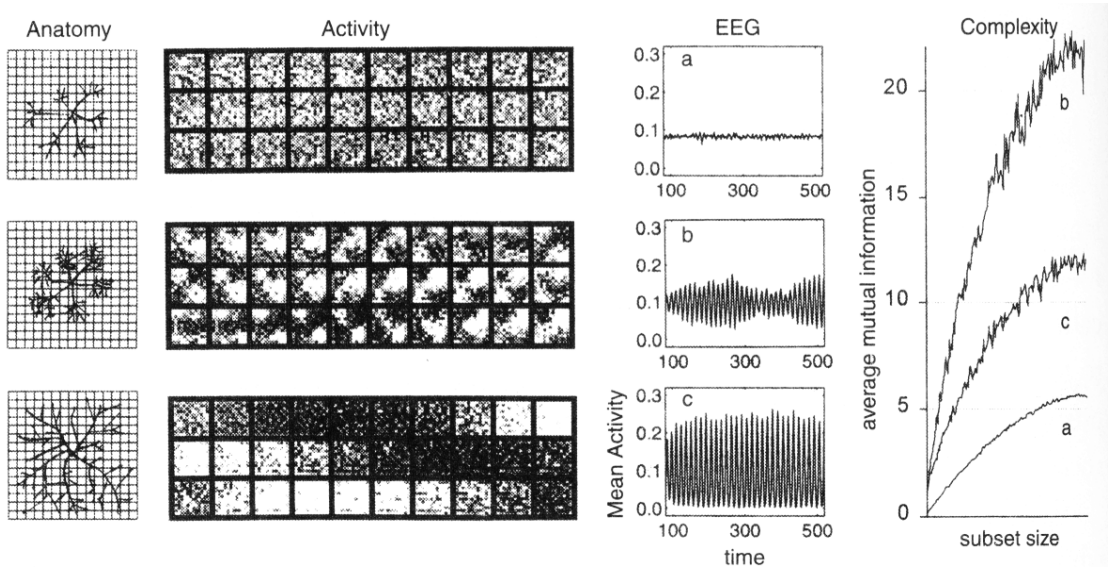
Dokumentuje to simulovanými experimenty s neuronovými sítěmi, jejichž struktura odpovídá anatomii vizuální mozkové kůry koček různého stáří. Staří jedinci mají v kůře lokalizovaná specializovaná centra, která však kvůli degeneraci tkáně nejsou integrována. Výsledkem jsou vzorce aktivity, které mají mnoho stavů, které ale nejsou synchronizovány (tj. téměř žádné skupiny neuronů nejsou pravidelně aktivovány zároveň). Výsledkem je nízká míra komplexity. Mladí jedinci ještě nemají rozvinutá specializovaná centra, ale jejich kůra je silně integrována. To vede k silné synchronizaci (většina skupin neuronů je aktivována současně). Vzhledem k malému počtu stavů je komplexita nízká. Konečně kůra normálního dospělého jedince obsahuje lokalizovaná specializovaná centra, která jsou dobře integrovaná. Vzorce aktivity mají mnohem více stavů než u mladého a méně než u starého jedince. Míra synchronizace je naopak vyšší než u starého a nižší než u mladého jedince a navíc kolísá, což znamená, že různě velké skupiny neuronů jsou aktivovány současně v různých časech. Komplexita je vysoká. [18]

Tononi [14] definuje adaptační komplexitu (matching komplexity), která neměří jen komplexitu interakcí uvnitř mozku, ale i komplexitu interakcí mezi mozkem a nějakou vrstvou vstupů. Tato komplexita je definována v podstatě analogicky neurální komplexitě a z formálního hlediska vcelku nic nového nepřináší (můžeme si představit, že vrstva vstupů je jistou specifickou součástí mozku).

Velmi zajímavá je však diskuse na konci uvedeného článku. Standardní aplikace informační teorie se většinou snaží řešit přenos signálů kanálem s omezenou kapacitou a šumem a využívají k tomu redukci dimenzí signálu (méně vstupů než výstupů), redukci redundance (komprese informace) a redukci šumu. V mozku však dochází k přesnému opaku těchto jevů. Počet jednotek mozku je mnohem vyšší než počet jednotek vstupních vrstev (zvyšování dimenze). V mozku probíhá paralelně mnoho silně korelovaných

[18] Pro lepší pochopení těchto procesů můžeme použít metaforu: Aktivita neuronů u starého jedince připomíná hluk v hostinci, kde se jednotlivé skupinky u stolů navzájem neposlouchají, aktivita u mladého jedince zase připomíná sjezd nějaké strany, kde se střídá projev jediného řečníka s hromovým aplausem. Konečně aktivita u dospělého jedince připomíná debatu několika lidí, nedá se určit jak bude kdo mluvit dlouho, může tam být i moderátor, který jim občas skočí do řeči, ale v podstatě si všichni vzájemně naslouchají.

procesů (vysoká redundance). Konečně signály přicházející ze vstupů jsou jen malou částí všech signálů, které mozek zpracovává, většina jich přichází ze zpětnovazebných (reentrant) signálů uvnitř samotného mozku (vysoký „šum“ vzhledem ke vstupnímu signálu). Tyto vlastnosti umožňují mozku vytvářet celou řadů možných variant reakcí na ten samý podnět a vybírat z nich v závislosti na kontextu, popřípadě „experimentovat“ s různými stejně vhodnými reakcemi na ten samý podnět. Adaptační komplexita umožňuje měřit citlivost vstupního podnětu na jeho kontext, který představují signály, které již v mozku probíhají, a rozlišovat tak podněty, které mají význam (kontext) od těch, které ho nemají. Autoři předpovídají, že vyšší adaptační komplexitu nalezneme u podnětů vedoucích ke konstruktivní percepci, jako jsou například různé smyslové klamy, vytváření třírozměrných objektů z dvourozměrných (např. Neckerova krychle), odlišování figury a pozadí a rozlišování smysluplných slov od nesmyslných.



Obrázek 4. Změna komplexity v závislosti na neuroanatomii [6]. První řádek se týká starého, druhý normálního dospělého a třetí mladého jedince. První sloupec zobrazuje anatomii kůry. Druhý sekvenci vzorců neurální aktivity (od levého horního k pravému dolnímu rohu). Třetí sloupec obsahuje průběh křivky EEG, jejíž amplituda je mírou synchronizace vzruchů. Ve čtvrtém sloupci je graf s křivkami průměrné vzájemné informace pro různé velikosti částí systému, komplexita je dána plochou pod křivkami.

Všechny tyto schopnosti korespondují s tím, jak jsme definovali samoorganizujícího se kognitivního agenta. Udržování komplexity vnitřní organizace agenta je umožněno (a zároveň umožňuje) selektováním změn prostředí (podnětů) důležitých pro jeho přežití (kontext) a vytváření neutrálních variant reakcí na podněty, které slouží k adaptaci.

3.2 Vztah mezi neurální komplexitou a kolísáním entropie

Existuje nějaká souvislost mezi neurální komplexitou a komplexitou definovanou v minulé kapitole, jejímž znakem bylo kolísání entropie? Všimněme si kolísání amplitudy EEG u vzorců neuronové aktivity s nejvyšší komplexitou. Souvisí toto kolísání nějak s kolísáním entropie?

Připomeňme, že kolísání entropie u komplexně se chovajících jednorozměrných celulárních automatů bylo způsobeno interakcemi lokalizovaných struktur. Tyto lokalizované struktury se objevují i u vícerozměrných celulárních automatů.

Obrázek 5. Příklad vývoje samoorganizujícího se dvourozměrného celulárního automatu [16]. Z počátečního náhodného vzorce buněk (první tři obrázky zleva zobrazují systém v prvních několika krocích) se postupně vytváří lokalizované struktury (dva obrázky vpravo zobrazují systém po několika stech krocích) připomínající vzorce neuronových aktivit s vysokou neurální komplexitou.

Anatomická struktura mozku na úrovni neuronů ovšem neodpovídá těm typům celulárních automatů, kde stav buňky v dalším kroku závisí pouze na několika málo jejích lokálních sousedech. Stav neuronů v mozku (když pomineme jiné faktory a omezíme se pouze na neurony), může záviset na velkém množství neuronů, jak z lokálního centra, tak ze vzdálenějších oblastí.

K zajištění anatomické podobnosti celulárního automatu a mozku můžeme buď považovat lokalizované struktury pohybující se v celulárním automatu za analogie signálů mezi lokálními centry v mozkové kůře. Nebo (a to je přesnější) můžeme definovat pravidla, která zajistí, aby stav buňky automatu závisel i na nějakém (typicky nízkém) počtu buněk vzdálenějších. V tom případě už to ale nebude klasický celulární automat, začne spíše připomínat neuronovou síť, ovšem míru kolísání entropie můžeme na takových automatech zkoumat stejným způsobem.

Entropie, kterou používá Tononi k výpočtu komplexity, je ovšem jiná než entropie, kterou zkoumal Wuensche. Tononi měří entropii vývoje stavů jedné části zkoumaného systému, kdežto Wuensche měří vývoj entropie systému rozčleněného na části jedné konkrétní velikosti. Tononi měří entropii dynamiky, kdežto Wuensche dynamiku entropie.

Abychom lépe pochopili vztah těchto dvou přístupů ke komplexitě, musíme analyzovat vztah amplitudy EEG a vývoje entropie (ve Wuenschově podání).

Amplituda EEG, která je mírou synchronizace v daném stavu, je pozitivně závislá na počtu současně aktivních skupin neuronů. Entropie jakožto míra neuspořádanosti, je nejvyšší právě tehdy, když je současně aktivní zhruba polovina neuronů (existuje nejvíce možných stavů - vzorců jejich aktivity). Entropie je nejnižší, když nejsou aktivní žádné nebo naopak, když jsou aktivní všechny neurony (existuje jen jeden stav, jeden možný vzorec aktivity). Jak je vidět z průběhu vzorců neuronové aktivity, u starého jedince je aktivní stále zhruba polovina neuronů, tj. entropie je vysoká a příliš se nemění (zde by mohla být analogie s náhodným chováním celulárního automatu). U mladého jedince se naopak pravidelně střídají stavy, ve kterých není aktivní téměř žádný neuron a stavy ve kterých jsou aktivní téměř všechny neurony, entropie v těchto stavech bude nízká, avšak při přechodu mezi nimi bude vysoká. Zde tedy již porovnání s uspořádanými celulárními automaty pokulhává, mohlo by však jít o analogii Wolframovy 2. třídy chování celulárních automatů s repetitivním chováním. U normálního dospělého jedince počet současně aktivních neuronů kolísá nepravidelně kolem poloviny (kdy je entropie nejvyšší), přičemž jen zřídka nastává stav, v němž je počet aktivních neuronů velmi nízký nebo velmi vysoký (entropie nejnižší). Entropie bude kolísat, přičemž velikost jejího kolísání bude proměnlivá. Tedy pokud budeme entropii sledovat v čase, bude se měnit její rozptyl (počítaný vždy z nějakého časového okna dané délky).

Souvislost mezi neurální komplexitou a kolísáním amplitudy EEG tedy můžeme převést na souvislost mezi neurální komplexitou a kolísání rozptylu entropie struktury vzorce neurálních aktivit.

Je možné ptát se ještě přesněji, jakou strukturu budou mít změny této entropie, popřípadě změny jejího rozptylu? Speciálně, jaká bude pravděpodobnostní distribuce těchto změn?

Vodítkem ke zodpovězení této otázky je zkoumání pravidel, která určují vývoj systému. V případě celulárního automatu jde o pravidla určující, na kterých buňkách a jakým způsobem bude záviset stav buňky v příštím kroku. Je možné zkoumat různá pravidla komplexního chování, měřit neurální komplexitu (resp. její ekvivalent pro celulární automaty) a porovnávat ji s pravděpodobnostním rozložením změn entropie. Těmto kvantitativním metodám ani analytickému řešení se zde věnovat nebudeme, můžeme však zformulovat hypotézu, že pravěpodobnostní distribuce změn entropie pro chování s nejvyšší neurální komplexitou je aproximováno mocninnýn zákonem (power law).

Mocninné zákony, které jsou další používanou charakteristikou komplexity, budeme zkoumat v následující kapitole.

4 Mocninné zákony

4.1 Bezškálové sítě

V této kapitole budeme zkoumat další běžně užívanou charakteristiku komplexity systému. Tou je podléhání vlastností systému mocninným zákonům. Nejprve budeme zkoumat prostorové uspořádání systémů a následně jejich dynamiku.

Anatomická a fyziologická struktura mozku, která vede k největší neurální komplexitě, je zároveň dobře integrovaná a diferencovaná. To je důsledek existence velmi silně lokálně propojených center, která jsou však dobře integrována, tedy existují mezi nimi dostatečně časté spoje na relativně dlouhé vzdálenosti.

Pro ilustraci uveďme schéma propojení specializovaných center v mozku makaka. Tato struktura má z hlediska teorie grafů mnohé zajímavé vlastnosti. Propojení na velkou vzdálenost vede k tzv. jevu „malého světa“ (small world), což znamená, že průměrná vzdálenost mezi dvěma uzly roste s velikostí grafu pouze logaritmicky[19].

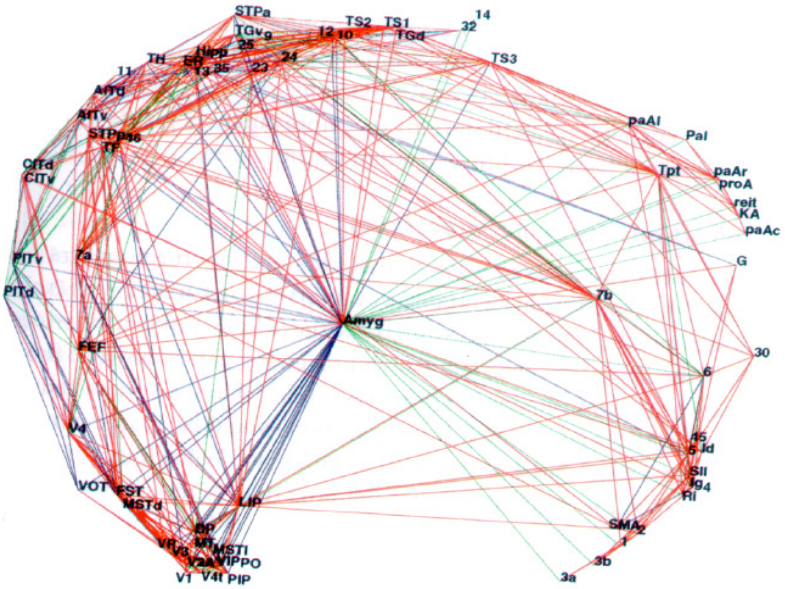
Přinejmenším některé studie např. [7] se snaží ukázat, že síť reprezentující spojení v mozku je jedním z typů sítí malého světa, tzv. bezškálovou sítí[20]. V této síti, na rozdíl od náhodných grafů, neexistuje typická konektivita[21], kterou zde můžeme chápat jako

[19] Podobný jev nastává i v náhodných grafech, kde ale chybí lokální shluky typické pro sítě malého světa. V náhodných grafech jsou každé dva uzly propojeny s jistou shodnou pravděpodobností. Pravděpodobnostní distribuce konektivit uzlů bude odpovídat Poissonovu rozdělení, s maximem pro nějakou typickou konektivitu, která bude odpovídat podílu počtu hran a počtu uzlů. Pro k větší než je tato typická konektivita klesá pravděpodobnost výskytu uzlu s konektivitou k exponenciálně (tj. velmi prudce pro velká k): $P(k) = a^k$, kde a je konstanta menší než jedna.

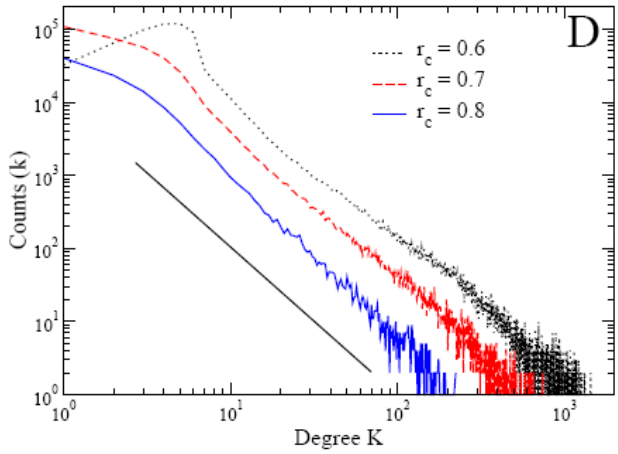
[20] Jistěže ne všechny sítě vztahů v mozku musí být bezškálové a splňovat mocninné zákony. Z hlediska adaptace se dá poukázat na to, že bezškálové sítě jsou sice odolné (robustní) proti náhodným poškozením, ale naopak velice citlivé proti cíleným poškozením uzlů s vysokou konektivitou. Naproti tomu náhodné sítě, nebo sítě, které jsou někde mezi náhodnými a bezškálovými sítěmi vykazují mnohem stabilnější odolnost i proti cíleným zásahům (existuje méně uzlů, na kterých závisí funkčnost celé sítě).

[21] Konektivita uzlu je počet uzlů se kterými je uzel spojen. Pravděpodobnostní distribuce konektivit uzlů je rozložení pravděpodobností, že náhodně vybraný uzel bude mít konektivitu k.

zobecněnou škálu, měřítko. Proto ona bezškálovost. V bezškálové síti je pravděpodobnost, že uzel bude mít konektivitu *k,* dána mocninným zákonem $P(k)=k^d$. Bližší vysvětlení je možno najít v Dodatku 1.



Obrázek 6. Schéma propojení korových center mozku makaka. Vzdálenost center ve schématu neodpovídá vzdálenosti center v mozku, ale míře jejich vzájemné propojenosti. Do centra byla pro přehlednost umístěna amygdala, která je propojena s většinou ostatních center [18].



Obrázek 7. Závislost mezi konektivitou uzlů a jejich počtem v síti odvozené z korelací aktivit mezi pravidelnými trojrozměrnými oblastmi (voxely) mozku. V této síti uzly odpovídají voxelům a hrany překročení hodnoty korelace r_c mezi nimi. Aktivita mozku byla snímána pomocí funkční magnetické rezonance (fMRI) [7].

Mocninná pravidla se většinou zobrazují na takzvaném log-log grafu, jehož osy jsou v logaritmickém měřítku, mocninná závislost má pak tvar přímky se sklonem, který odpovídá exponentu *d*. Příklady těchto sítí nacházíme všude kolem nás. Jde o systémy, které většinou vznikají postupně, jsou složené z mnoha vzájemně interagujících částí a musí optimalizovat rychlost vzájemné komunikace na větší vzdálenosti. Příkladem

takových sítí jsou sítě sociálních vztahů, internet a struktura sítě WWW, síť vztahů mezi slovy, síť citačních odkazů a mnoho dalších[22].

V biologii je natolik častý typ struktur, jejichž charakteristiky podléhají mocninným zákonům, že pro ně existuje speciální termín „allometrie“. Nemusí se ovšem jednat jen o bezškálové sítě. Například mezi vztahy veličin charakterizující mozek savců, které podléhají mocninným zákonům, patří vztahy mezi objemem šedé a bílé hmoty, hustotou neuronů a délkou a průměrem axonů [9].

4.2 Samoorganizované kritično

Bezškálovost však není typická jen pro prostorové uspořádání samoorganizujících se systémů. Statistická fyzika popisuje řadu procesů, které je možné charakterizovat fraktální, bezškálovou strukturou a vývojem podléhajícím mocninným zákonům. Jde zejména o fázové přechody, tedy stavy, kdy jsou fyzikální parametry (typicky teplota a tlak) nastaveny tak, že není možné rozlišit mezi dvěma stavy hmoty. Tento stav se nazývá kritický (critical). Kritické stavy jsou v případě fázových přechodů většinou velice nestabilní a závislé na velmi přesném nastavení parametrů, dlouhodobě jsou udržitelné pouze v laboratoři.

Bak a jeho kolegové [2,3] ukázali, že existuje celá řada systémů, které mají jako svůj atraktor právě kritické stavy. Přirozeně a robustně se ke kritickému stavu vyvíjejí, pokud jsou z něho vyvedeny nějakou vnější či vnitřní fluktuací zase se do něj navrací. Dynamiku těchto systémů nazval Bak samoorganizované kritično (self-organised criticality)[23].

Příkladem takového systému je hromada písku, na kterou postupně přisypáváme zrníčka. Hromada se postupně zformuje do kužele, jehož stěny budou mít stálý sklon. Přisypané zrníčko nemusí způsobit téměř žádné změny, ale může strhnout i rozsáhlou lavinu (avalanche)[24].

V [3] je popsána celá řada systémů, které je možné popsat v termínech samoorganizovaného kritična. Patří mezi ně dynamika zemětřesení, povodní, dopravních systémů, vývoje finančních trhů nebo evoluce druhů.

---

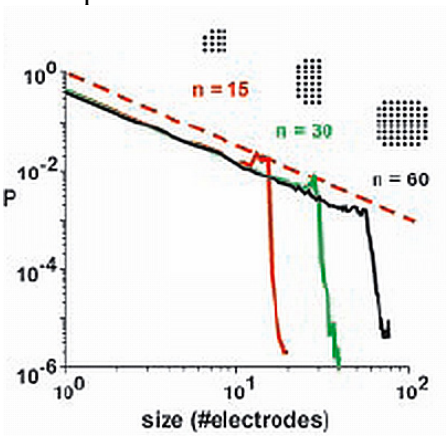
[22] Dobrý model pro vznik bezškálové sítě popisuje Barabási (1999). Síť vzniká postupným přidáváním uzlů k síti, přičemž s vyšší pravděpodobnost přidáme nový uzel k již existujícímu uzlu s vyšší konektivitou. Tento princip hezky popisuje úsloví „Rich get richer.“ a je to mj. příklad samoorganizujícího se procesu. Zajímavé by bylo zkoumat platnost tohoto principu například v ontogenetické dynamice růstu neuronů a synapsí mezi nimi.

[23] Terminologická poznámka: Někteří autoři zabývající se samoorganizací a komplexitou používají termínu self-organized complexity jako synonymum self-organized criticality. Tyto termíny předpokládají existenci samoorganizace, která nevede ke komplexnímu chování, ale buď k uniformnímu, repetitivnímu, nebo naopak náhodnému chování. V tomto článku jsme ale definovali samoorganizaci přímo jako proces udržování či zvyšování komplexity. Takto definovaná samoorganizace je pak jedním z typů chování tzv. komplexních systémů (viz. pozn 1), mezi něž patří i procesy vedoucí k uniformnímu, repetitivnímu či náhodnému chování.

[24] Bak modeloval hromadu písku pomocí jednoduchého celulárního automatu. Přestože se Bak snaží dokázat, že chování tohoto idealizovaného modelu odpovídá reálným procesům, objevily se i protipříklady, že reálné hromady zemin mají jiné vlastnosti (díky tomu, že v nich působí mnoho vnějších faktorů jako je vlhkost, nepravidelnost a nestejnorodost zrn apod.). To však neubírá Bakovu modelu na konceptuální názornosti a přesvědčivosti.

Pravděpodobnostní distribuce velikosti změn v těchto systémech je dána mocninným zákonem. To v praxi znamená, že v systému bude docházet ke změnám úměrným časovému měřítku. Neexistuje nějaká typická velikost změn (analogie neexistence typické konektivity uzlu v bezškálových sítích) a bude tedy docházet i k velice rozsáhlým změnám. Tyto rozsáhlé změny jsou často trefně nazývány katastrofy, či laviny (i mimo původní kontext těchto pojmů). I když výskyt katastrof není pravidelný, jejich počet v daném časovém úseku bude dán stabilně mocninným zákonem. Například za jedno tisíciletí přichází deset tzv. stoletých povodní, ovšem to neznamená, že stoletá povodeň přichází jednou za sto let, povodně mohou být nahromaděné v nějakém mnohem kratším časovém úseku.

Výzkum samoorganizované kritické dynamiky kognitivních procesů je na nejnižších úrovních kognice dosud v počátcích, ale do budoucna vykazuje velký potenciál pro hlubší pochopení mezi strukturou kognitivních agentů a jejich dynamikou. Jedna z nejcitovanějších studií v této oblasti [5] ukazuje výskyt samoorganizovaných vzorců neuronové aktivity v řezech krysí korové tkáně. Tyto řezy byly buď uměle pěstované z kultur neuronálních buněk nebo to byly čerstvě odebrané (acute) preparáty z krysích mozků. Aktivita neuronů byla snímána maticí elektrod. V obou případech bylo prokázáno, že intenzita lavin neuronových vzruchů (vyjádřená jako počet současně aktivních elektrod) podléhá mocninnému zákonu s exponentem zhruba -1.5.



Obrázek 8. Aktivity elektrod snímající řez uměle pěstovanou krysí korovou tkání [5]. Pravděpodobnost výskytu stavů při nichž byl současně aktivní jistý počet elektrod klesá podle mocninného zákona s exponentem zhruba -1.5 . Různé odstíny čar představují různě velké použité matice elektrod. Závislost se prudce láme před dosažením maximálního počtu elektrod.

Do budoucna, s rozvojem technik pro přesnější (zejména neinvazivní) zkoumání mozku, je možné očekávat nová data o vzorcích neuronové aktivity i ze živých zvířat a lidí.

Vztahu mezi mocninnými zákony a vyššími kognitivními funkcemi jako je pozornost a vnímání smyslových podnětů, získávání dovedností, učení, vybavování a zapomínání, se věnuje rozsáhlá psychologická literatura. Téměř všudypřítomná existence mocninných zákonů v těchto oblastech (přičemž ovšem existují podstatné výjimky) bývá někdy zpochybňována, jako metodologický artefakt či explanační konstrukt [1].

Tyto psychologické studie se však nepokoušejí vysvětlovat vztah mezi mocninnými zákony a vyššími kognitivními funkcemi pomocí samoorganizace.

Naopak existuje řada pokusů modelovat kognitivní agenty pomocí konekcionistických systémů (jako jsou například celulární automaty a umělé neuronové sítě spolu s umělou evolucí). Jak jsme ukázali v předchozích kapitolách, pro tyto systémy jsou definované míry komplexity (neurální komplexita) nebo alespoň znaky komplexity (kolísání entropie) založené na pojmech z teorie informace. Jak ukazuje Dodatek 2, mezi charakterizováním komplexity pomocí mocninných zákonů a pojmy teorie informace existuje hluboká souvislost, která dává vodítka pro další studium komplexity.

5 Závěr

Samoorganizující se systém můžeme metodicky chápat jako systém, jehož atraktory tvoří ve stavovém prostoru trajektorie s kritickou, resp. obecněji komplexní dynamikou. Tato definice odpovídá tomu, jak jsme definovali aktivitu (kognici) autonomního, situovaného a vtěleného kognitivního agenta. Tento agent se chová takovým způsobem, který umožňuje udržení či zvýšení komplexity jeho vnitřní organizace. Organizaci jsme definovali jako strukturu prostorových vztahů, dynamických interakcí a vztahů mezi nimi. Samoorganizující se kognitivní agent tedy nemusí být nutně biologický. Může to být jakýkoliv systém (například sociální, či uměle vytvořený), který má výše uvedené vlastnosti.

Samoorganizace je natolik široký koncept, že může sloužit jako metodický rámec pro popis kognice jako aktivní schopnosti agenta rozpoznávat změny prostředí významné pro jeho přežití, reagovat na ně, adaptovat se na ně a ovlivňovat prostředí.

Na druhé straně vzniká obava, zda koncept samoorganizace není bezbřehý a může nám přinést konkrétní vodítko pro analýzu empirických dat, tvorbu kvantitativních modelů a konstrukci analytických pravidel popisujících kognici.

Klíčový se ukazuje pojem komplexity (nebo méně obecný pojem kritična), který je v současné době charakterizován (alespoň v teorii samoorganizace) třemi formálně dobře definovanými vlastnostmi. Nejvágnější je kolísání entropie (či rozptyl entropie) po sobě následujících stavů systému. Poměrně dobře definovanou charakteristikou je neurální komplexita, založená na vzájemné informaci mezi vyvíjejícími se částmi systému. Třetí jsou mocninné zákony, která jsou příznakem bezškálovosti systému.

Otevírá se zde široké pole výzkumu. Zejména jde o zkoumání vztahů mezi výše uvedenými charakteristikami komplexity. Obecně je pak hlavním úkolem pochopení vztahů mezi prostorovým uspořádáním částí systému a dynamikou jejich interakcí.[25]

Dále jde o zpřesnění charakteristik komplexity, zejména v případe „všudypřítomných“ a proto jen málo informativních mocninných zákonů. Je možné, že existují vztahy, které nejsou mocninné, ale přitom charakterizují komplexitu lépe (například distribuce konektivit uzlů sítě, která není ani náhodným grafem, ani bezškálová, ale někde „mezi“). Podobně je nutné zpřesnit charakteristiku komplexity založenou na kolísání či rozptylu entropie. Zde je třeba studovat možné pravděpodobnostní distribuce změn entropie a jejich vztah k neurální entropii.

Existují tři hlavní metody, které přispívají k vhledu do těchto souvislostí a které je nutno transdisciplinárně propojovat. Jde jednak o kvantitativní modely či simulace,

[25] Zde je nutné případné vážnější zájemce o tuto problematiku upozornit, že v tomto článku je, z důvodů omezeného rozsahu, vynechána rozsáhlá teorie dynamických systémů, využívající formalismů vlastních hodnot, vektorů a funkcí systému

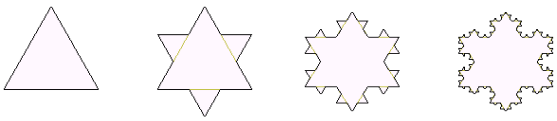
získávaní a zpracování empirických dat a matematicko - analytická řešení[26]. Snad k takovému propojení přispěl i tento článek.

Poděkování: Vznik tohoto článku byl umožněn díky grantu GA ČR č. 401/03/H047.

Dodatek 1.

Tento dodatek objasňuje v krátkosti pojem bezškálovosti.

Obecně bezškálovost znamená, že se pravidla popisující vlastnosti systému nemění (nebo jsou soběpodobná) se změnou měřítka (škály). Triviální příklad jsou tradiční geometrické objekty, vztah mezi stranou čtverce a jeho obsahem bude vždy stejný bez ohledu na měřítko a bude vyjádřen druhou mocninou, což odpovídá faktu, že čtverec má dimenzi 2. Podobný vztah je mezi stranou krychle a jejím objemem, zde bude tento vztah vyjádřen třetí mocninou, krychle má dimenzi 3. Úsečka má dimenzi 1. Tyto vztahy lze zobecnit na neceločíselné dimenze, které popisují vlastnosti tzv. fraktálních objektů. Změna měřítka u fraktálního objektu vede k růstu jeho zobecněného objemu[27], který je vyjádřen $N(r) = r^d$ (všimněme si, že jde o mocninný zákon), kde r je měřítko, $N(r)$ můžeme chápat jako počet stavů, které sytém může při daném měřítku nabývat (nebo počet zobecněných jednotkových objemů, které jsou potřeba k jeho pokrytí) a d je fraktální dimenze. Vidíme, že pro výše uvedené příklady úsečky, čtverce a krychle (d = 1, 2, 3) tento vztah funguje. Jedním z nejznámějších a nejilustrativnějších příkladů fraktálního objektu je Kochova vločka.



Obrázek 9. Kochova vločky pro čtyři změny měřítka

S každou trojnásobnou změnou měřítka (až do nekonečna) vzniká útvar mající čtyřnásobný počet stavů (čtyřnásobný zobecněný objem), v tomto případě čtyři v ploše poskládané úsečky. Platí tedy $4=3^d$ a fraktální dimenze Kochovy vločky je tedy $d= \log_3 4 = \log 4/ \log 3$, tj. zhruba 1.26. Obecně pak platí, že dimenze nějakého systému relativní k měřítku r je $d= \log_r N(r)$.

Fraktální objekty si nemusí být na různých škálách zcela podobné jako Kochova vločka, jejich části mohou být organizovány náhodně, musí však splňovat podmínku, že d je stejné pro všechna měřítka [28].

[26] Je pravděpodobné, že podobně jako v případě fraktálů, existují již mnohé vhodné matematické struktury, které však zatím nejsou dostatečně aplikovány.

[27] Zobecněný objem je objem v abstraktním prostoru s nějakou obecnou (i ne-Euklidovskou) metrikou.

[28] Existuje celá řada formálních přístupů, které zpřesňují pojem fraktální dimenze. Zde jsme se při vysvětlování opřeli o dvě dle mého názoru nejintuitivnější definice. Jednak částečně o tzv. entropickou či pokrývací dimenzi (pokrytí zobecněného objemu pomocí jednotkových objemů) a homotetickou dimenzí (počet nově vzniklých útvarů při změně měřítka). Homotetická dimenze je velmi jednoduchá, pro některé útvary bychom však dostali takové kombinace r a $N(r)$, že by výsledek vedl k nesprávnému či nekonečnému d. Dokonalejší definice fraktální dimenze (jako je výše zmíněná pokrývací nebo Hausdorffova-Besicovitchova dimenze) analyzují, která d dávají konečné výsledky pro různé limity r. Blíže viz (Mandelbrot, 2003)

V reálných systémech samozřejmě existují omezení. Největší měřítko je dané velikostí zkoumaného systému, nejmenší měřítko pak rozlišovací schopností pozorovacích přístrojů či modelu systému.

Když se vrátíme k bezškálovým sítím, můžeme si představit, že uzly jsou hierarchicky uspořádané podle konektivity, hodnota konektivity k odpovídá měřítku r a zobecněný objem systému $N(k)$ odpovídá počtu uzlů s konektivitou k (pravděpodobnost $P(k)$ dostaneme jako podíl $N(k)$ a celkového počtu uzlů sítě). Pro všechna k pak, jak už jsme zmínili, platí $P(k) = k^d$. Dimenze d se u většiny reálných bezškálových sítí pohybuje mezi -1 a -3 (kladné hodnoty dimenze dostaneme pokud bychom jako měřítko brali 1/k).

Dodatek 2.

Tento dodatek objasňuje v krátkosti vztah mezi pojmem dimenze a entropie.

Jak jsme ukázali v Dodatku 1, dimenze systému relativní k měřítku je určena vztahem: $d= \log_r N(r)= 1/\ln(r) * \ln(N(r))$.

Tento vztah odpovídá rovnici termodynamické entropie: $S = k \ln W$.
Zde S je entropie, k je Boltzmanova konstanta, a W počet mikroskopických stavů, které jsou slučitelné s makroskopickými charakteristikami systému. Zatím jsme předpokládali, že pravděpodobnost všech stavů systému je stejná, $p_i = 1/N(r)$. Platí:

$$d = \log_r N(r) = - \log_r 1/N(r) = - \log_r ((1/N(r))^{1/N(r)})^{N(r)} = -\Sigma_{i=1..N(r)} 1/N(r) \log_r 1/N(r)$$

Informační entropie $H(X)$, je analogií termodynamické entropie, pouze počítáme s tím, že stavy mohou mít různou pravděpodobnost p_i a místo o systému mluvíme většinou o pravděpodobnostní veličině (například X). Platí nerovnost:

$$- \Sigma_{i=1..N(r)} 1/N(r) \log_r 1/N(r) \geq - \Sigma_{i=1..N(r)} p_i \log_r p_i = H(X)$$

Tato nerovnost vyjadřuje nám již známý fakt, že pro daný počet stavů je entropie nejvyšší v případě, že všechny stavy pravděpodobnostní veličiny X jsou stejně pravděpodobné.

Jak dimenze tak entropie vyjadřují tedy mocninný vztah mezi jednotkou měření a stavy, které systém může nabývat.

První důležitý rozdíl mezi nimi je postavení závislých a nezávislých proměnných a konstant. V případě informační entropie máme většinou k dispozici nějaké pevně dané minimální měřítko, typicky velikosti $r=2$, tj. bit. Nezkoumáme jak se budou měnit stavy systému se změnou měřítka, předpokládáme, že jsou na měřítku nezávislé. Stavy systému (resp. jejich pravděpodobnostní distribuce) jsou nezávislou proměnnou, entropie je závislou proměnnou, měřítko je konstantní. V případě dimenze je naopak dimenze konstantní (pokud zkoumáme bezškálový systém), nezávislá proměnná je měřítko a závislá počet stavů systému.

Druhý rozdíl je, že v případě dimenze (alespoň pokud jako dimenzi interpretujeme exponent mocninného zákona) se většinou neuvažuje o pravděpodobnostní distribuci stavů systému. Pokud bychom entropii chápali jako obecnější verzi dimenze, i pro systémy s pravděpodobností distribucí stavů, mohli bychom se ptát po vlastnostech systémů, které mají stejnou entropii pro různá měřítka (zde musíme ovšem opustit předpoklad, že se stavy a jejich pravděpodobnostní distribuce se změnou měřítka nemění).

Zajímavou ukázkou, kam nás takové tázání může dovést je hlubší porozumění vztahu mezi bezškálovostí a vzájemnou informací mezi částmi a celkem systému (a tudíž i mezi bezškálovostí a neurální komplexitou). Vzájemná informace mezi částí systému velikosti k a celkem systému je nejvyšší v případě, že se rovnají entropie celku systému $H(X)$, entropie částí systému $H(X_j^k)$ i entropie zbytku systému $H(X - X_j^k)$. Neurální

komplexita sčítá průměrné vzájemné informace pro všechny velikosti částí systému a bude nejvyšší pro dané $H(X)$, pokud bude výše uvedená rovnost platit pro všechny velikosti (měřítka) k. Takový systém je z hlediska k bezškálový. Jeho dimenzí je entropie $H(X) = H(X_j^k) = H(X - X_j^k)$. Většinou jde o triviální případy vysoce uspořádaných systémů s nízkým počtem stavů. Lze ale zkoumat, zda se komplexní systémy nevyznačují nějakou typickou pravděpodobností distribucí stavů systému, která maximalizuje neurální komplexitu tím, že umožňuje, aby byl systém bezškálový či bezškálovosti blízký.

Literatura

[1] Anderson, R. B.: The power law as an emergent property. *Memory and Cognition* **29**(7) (2001) 1061-1068.

[2] Bak, P., Tang, C. and Wiesenfeld, K.: Self-organized criticality. *Phys. Rev.* **A 38** (1988) 364–374.

[3] Bak, P.: *How Nature Works: The Science of Self-Organized Criticality.* New York: Copernicus 1996.

[4] Barabási, A., Reka, A.: Emergence of scaling in random networks. *Science* **286** (1999) 509-512.

[5] Beggs, J., Plenz, D.: Neuronal Avalanches in Neocortical Circuits. *Journal of Neuroscience* **23**(35) (2003) 11167–11177.

[6] Edelman, G., Tononi, G.: *Universe of consciousness*. New York: Basic Book 2000.

[7] Eguíluz, V. M., Chiavalo, D. R., Cecchi, G. A., Baliki, M., Apkarian, V. A.: Scale-Free Brain Functional Network. *Physical Review Letters* **94** (2005) 018102.

[8] Gell-Mann, M.: What Is Complexity? *Complexity* **1** (1995) 16-19.

[9] Harrison, H. K., Hof, P. R., and Wang, S. S.: Scaling laws in the mammalian neocortex: Does form provide clues to function? *Journal of Neurocytology* **31** (2002) 289–298.

[10] Kauffman, S.: *At Home in the Universe: The Search for the Laws of Self-Organization and Complexit*. Oxford: Oxford University Press 1995.

[11] Kováč, L.: Komentovaný úvod do kognitívnej biológie. In: Kvasnička, V., Kelemen, J. (eds.). *Kognice a umělý život IV*. Opava: Slezská univerzita 2004.

[12] Mandelbrot, B.: *Fraktály*. Praha: Mladá fronta 2003.

[13] Maturana, H., Varela, F.: *Autopoiesis and Cognition: The Realization of the Living*. Boston: D. Reidel 1980.

[14] Tononi, G., Sporns, O., and Edelman, G.: A complexity measure for selective matching of signals by the brain. *Proc. Natl. Acad. Sci. USA*. **93** (1996) 3422-3427.

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

[16] Wolfram, S.: *A New Kind of Science*. Champaign, IL: Wolfram Media 2002.

[17] Wuensche, A.: Discrete Dynamical Networks and their Attractor Basins. In: *Complex Systems '98.* Sydney, Australia: University of New South Wales 1998.

[18] Young, L.: The organization of neural system in the cerebral cortex. *Proceedings of the Royal society* Series B, **252** (1993) 13-18.

Kognitívne východiská informačnej vedy

Mirka Grešková[1]

Abstrakt. Príspevok prináša konceptuálnu analýzu východiskových pojmov kognitívnej paradigmy informačnej vedy a kognitívnych vied. Jedným z cieľov je načrtnúť genézu kognitívnej paradigmy, ale aj jej kritiku a nájsť miesto informačnej vedy v kognitívnych vedách. Príspevok sa sústreďuje na analýzu koncepcií informačného správania, interakčného prístupu a príklady ich aplikácie v dizajne informačných systémov.

1 Úvod

Informačná veda sa zo všeobecného hľadiska zaoberá ľudskými aspektmi spracovania informácií. V najširšom slova zmysle sa zameriava na komunikáciu informácií v spoločnosti. Zasahuje do zásobníka rôznych vedných disciplín. O informačnej vede je možné hovoriť cez prizmu filozofie, sociológie, antropológie, etnológie, lingvistiky, psychológie, histórie, sociálnej komunikácie, ale aj umelej inteligencie, neurofyziológie, biológie, atď. Pred viac ako tromi dekádami informačná veda nadväzuje dialóg s kognitívnymi vedami. Boli tak položené základy kognitívnej paradigmy, ktorá predstavuje „obrat“ [31,33] v informačnej vede. Predstavuje odklon od systémového hľadiska, ktoré sa vyznačuje exaktnosťou, logickým kartezianizmom, technologickým prístupom a prísnymi kvantitatívnymi hľadiskami. Na javisko vstupuje človek, ktorý je v interakcii s informačným prostredím a svojím správaním sa adaptuje na okolie.

Cieľom príspevku je predstaviť informačnú vedu v kontexte kognitívnych vied. Nasledujúca časť analyzuje konceptuálne východiská kognitívnej paradigmy informačnej vedy a kognitívnych vied. Výsledný konceptuálny graf sumarizuje vzťahy medzi jednotlivými pojmami. Tretia časť sprehľadňuje genézu kognitívnej paradigmy, ale aj kritiku kognitívnych východísk informačnej vedy. Štvrtá časť približuje vzťah informačnej vedy a kognitívnych vied. Koncepcie kognitívnej paradigmy predstavuje piata časť, ktorá je ďalej členená na podčasti venované informačnému správaniu, interakčnému prístupu a aplikáciám kognitívnych vied v dizajne informačných systémov.

2 Konceptuálne východiská

Konceptuálna analýza je založená na charakteristike základných pojmov v oblastiach kognitívnej paradigmy informačnej vedy (pozri časť 2.1) a kognitívnych vied (pozri časť 2.2). Snažili sme sa doplniť a rozšíriť prehľad základných pojmov zverejnený v predchádzajúcich prácach [24,25,26]. Východiskové pojmy a vzťahy medzi nimi sú vizualizované v konceptuálnom grafe (pozri obr. 2).

2.1 Kognitívna paradigma informačnej vedy

Informačná veda operuje s fundamentálnym pojmom *informácie*, ktorý ohýba pri charakteristike predmetov záujmu teórie a výskumu. Existuje mnoho definícií informácie, ktoré vychádzajú

[1] Katedra knižničnej a informačnej vedy, Filozofická fakulta Univerzity Komenského, Šafárikovo nám. 6, 818 01 Bratislava, Slovensko, E-mail: mirka.greskova@fphil.uniba.sk

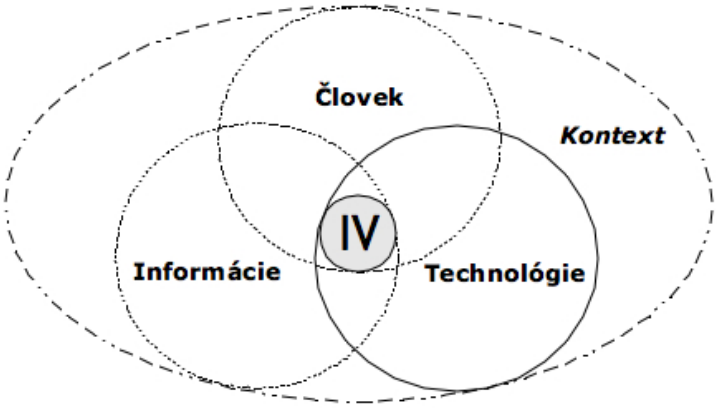
z rôznych hľadísk. Najčastejšie sa uvažuje o kvantitatívnom a kvalitatívnom rozmere informácie. Rozlišuje sa sémantická, syntaktická a pragmatická stránka informácie. Vo všeobecnosti sa dá zhrnúť, že definície informácie sa viažu zvyčajne na tri druhy fenoménov: kognitívny proces, odovzdávanie poznatkov a označovanie objektov [10]. V tradičnom (shannonovskom) ponímaní je informácia statickou pravdepodobnosťou určitého signálu/znaku na vstupe konkrétneho systému. Znižovaním pravdepodobnosti určitého znaku sa zvyšuje jeho informačná hodnota. Spracovaním signálu systémom sa znižuje úroveň entropie a systém sa dostáva do stavu vyššej miery usporiadanosti [57]. Odlišný prístup predstavuje definícia informácie ako tej, ktorá konštruuje významy reality v rôznych premenlivých formách [52]. Významy interpretovaných informácií vedú k zmene stavu poznania (transformáciám kognitívnych štruktúr).

Pojem informácie sa často zamieňa s dátami, poznatkom a znalosťou. *Dáta* predstavujú komunikované formalizované označenia, ktoré sa po vnímaní a interpretácii stávajú informáciami (sú teda potencionálnymi informáciami). *Znalosť* transformuje dáta na informácie a je výsledkom aktívneho učenia sa a porozumenia informácií [57]. Podľa Šlapáka [57] je možné znalosť vysvetliť v kontexte jej použitia. Znalosť je tvorená vzájomne poprepájanými štruktúrami poznatkov, ktoré sa dynamicky menia. *Poznatok* vytvára kognitívnu štruktúru, nad ktorou sa realizujú kognitívne procesy.

Rôzne pohľady formujú odlišné definície *informačnej vedy*. Podľa definície ASIS (American Society for Information Science, Americká spoločnosť pre informačnú vedu) z roku 1975 [30] sa informačná veda zaoberá vytváraním, zberom, organizáciou, interpretáciou, uchovávaním, prieskumom, rozptylom, transformáciou a využívaním informácií s dôrazom na aplikácie moderných technológií v týchto oblastiach. Hľadá spôsoby vytvárania a štruktúrovania vedeckých, technologických a systémových poznatkov súvisiacich s prenosom informácií. Túto definíciu možno označiť ako systémovo-orientovanú.

Veľmi široko poňal informačnú vedu Soergel [48] a v jej jadre poukázal na snahu o pochopenie používateľov, ktorí hľadajú významy a riešenia problémov a reprezentáciu poznatkových štruktúr na podporu konštrukcie významu a riešenia problémov.

Informačná veda dáva do vzťahu informácie, človeka a technológie v spoločenskom kontexte (obr. 1). V interakciách medzi jednotlivými prvkami sa realizujú rôznorodé procesy, ktoré prebiehajú simultánne, navzájom sa dopĺňajú a prelínajú.



Obrázok 1. Informačná veda (IV): človek, informácie a technológie.

Kognitívna paradigma informačnej vedy vychádza z kognitívnych vied a ohniskom záujmu sú psychologické aspekty spracovania informácií človekom [29]. Limbergová [39] predstavila konceptuálny model pozostávajúci z informácií uložených v informačných systémoch, zo sprostredkovateľov a používateľov. Tieto zložky vytvárajú triádu, ktorá predstavuje jadro štúdia informačnej vedy v kontexte kognitívnej paradigmy. V súčasnosti sa preferuje holistický pohľad, ktorý zdôrazňuje kontext a citlivo vníma situáciu človeka (ľudí) pri vyhľadávaní informácií a informačnom prieskume.

V kognitívnej paradigme sa vyvinula používateľská a technologická línia. Z aspektu používateľa je najviac rozoberaná problematika informačného správania a vyhľadávania informácií. Technologický smer napĺňa orientácia na výskum a tvorbu praktických systémov na reprezentáciu, organizáciu a vyhľadávanie informácií (digitálne knižnice, vyhľadávacie nástroje a pod.). Obe línie kognitívnej paradigmy sa prelínajú a stretávajú v spoločných priesečníkoch. Prepájajúcim prvkom je najmä interakčný prístup (pozri časť 5.2), ktorý harmonizuje tieto aspekty.

Skúmanie *informačného správania* dáva odpovede na otázky, ako sa človek adaptuje na informačné prostredie prostredníctvom vyhľadávania, spracovávania a využívania informácií, ako rieši problémy pomocou získavania informácií, aké postupy využíva, aké zložky správania sa u neho prejavujú, atď. Vďaka výskumom informačného správania je možné zužovať interakciu medzi človekom a strojom a zefektívňovať systémy, ktoré sú využívané ľuďmi.

2.2 Kognitívne vedy

Kognitívne vedy predstavujú oblasť, ktorá sa špecializuje na štúdium kognície v interdisciplinárnom rámci. V prieniku vymedzuje CSS (Cognitive Science Society, Spoločnosť kognitívnej vedy) [13] antropológiu, umelú inteligenciu, pedagogiku, lingvistiku, neurovedu, filozofiu a psychológiu. Základné objekty skúmania kognitívnych vied sú u rôznych autorov vnímané odlišne. Najčastejšie sa uvádza kognícia, kognitívne procesy (myslenie, inteligencia), mentálne javy a výpočty.

Východiskom kognitívnych vied je výpočtová koncepcia a hľadisko *spracovania informácií* [56]. Vysvetľovanie fungovania mysle a mozgu, pochopenia ich vzťahu je založené na počítačovej analógii. Za zakladateľov teórie spracovania informácií sa považujú Newell a Simon. Spracovanie informácií je transformačný proces, ktorý sa realizuje ako súbor procedúr kódovania, dekódovania, prenosu, kombinovania, uchovávania, vybavovania informácií [55]. Otvorená ostáva otázka správnosti využitia pojmu spracovanie informácií vo výpočtovom kontexte. V oblasti informatiky sa pojem informácie používa voľnejšie. Informačná veda je prísnejšia a odlišuje medzi spracovaním informácií u človeka a strojovým spracovaním dát/údajov.

Kognitívna psychológia je subdisciplínou psychológie a predstavuje jadro kognitívnych vied. Termín sa zaužíval v roku 1967 vďaka Neisserovi, ktorý publikoval monografiu s rovnomenným názvom. V jeho ponímaní kognitívna psychológia zdôrazňuje existenciu konceptuálnej štruktúry mysle. Dnes je objekt tejto disciplíny rozšírený o mentálne reprezentácie. Vzťah medzi konceptuálnou štruktúrou a mentálnou reprezentáciou je vysvetlený v ďalšom texte. Cieľom kognitívnej psychológie je odpovedať na otázky ako človek prijíma, spracúva a využíva informácie [45].

Kognícia súvisí s poznávaním vyjadreným v termínoch kognitívnych procesov [44]. Gál [22] kogníciu prepája s procesmi, dispozíciami a výsledkami spracovania

informácií mysľou. Stillings [56] ju stotožňuje s vnímaním a poznávaním. Kognícia je cieľom aj zmyslom prúdu života každého človeka. Práve preto stojí na piedestále kognitívnej psychológie a kognitívnych vied.

Kognitívne procesy sú, okrem ďalších štruktúr a reprezentácií, stredobodom záujmu kognitívnych vied. Patria medzi nich vnímanie, myslenie, pamäť, atď. Realizujú sa nad mentálnymi reprezentáciami objektov. *Mentálne reprezentácie* odzrkadľujú objekty sveta a vytvárajú obsah mysle. Sú vnútornými reprezentáciami vonkajších podnetov [55]. S termínom operuje výpočtová teória mysle v kontexte analógie mysle ako počítača. Thagard [58] medzi základné mentálne reprezentácie priraďuje logické predpoklady, pravidlá, pojmy, obrazy a analógie. Mentálne reprezentácie podľa neho využívajú mentálne procesy dedukcie, vyhľadávania, prieniku, rotácie a prieskumu.

Kognitívna štruktúra je typom mentálnej reprezentácie. Medzi prvými o nej hovorí Piaget v roku 1929. Jeden z prístupov k interpretácii kognitívnych štruktúr prezentujú *kognitívne mapy*. Ide o reprezentácie okolia z pohľadu priestorových vzťahov. Kognitívnou mapou je napr. predstava cesty z bodu A do bodu B, vďaka nej sa ju nemusíme stále nanovo učiť a dokážeme sa orientovať v priestore. V experimentálnom výskume sa kognitívnym mapám venovali Tolman, Honzik, Frisch, Thorndyke a Hayes-Roth [55].

Kognitívna štruktúra obsahuje *konceptuálnu štruktúru* [8,31,54], ktorá pozostáva z pojmov, pravidiel ich využívania a ich prepojení (vzťahov). *Pojem* je idea vytvorená v mysli, jednotka, ktorá reprezentuje znak (sprostredkovaný vnímaním) a objekt sveta (sprostredkovaný skúsenosťou) v mysli človeka. Pojem môže mať funkciu kategórie, a organizovať tak ďalšie pojmy, ktoré majú spoločné charakteristiky.

Schémy a scenáre predstavujú typy konceptuálnej štruktúry. Sú komplexnejšie ako pojmy. Schémy sú vnútorne poprepájané vzťahmi, ktoré reprezentujú v pamäti človeka príbehy. Časti schémy môžu byť stabilné alebo premenlivé. Napr. schéma „idem do kina“ môže pozostávať zo stabilných častí ako zakúpenie lístka, usadenie sa na miesto a premenlivých častí ako názov filmu a pod. Špeciálnym prípadom schémy je scenár. Popisuje následnosť často opakovaných udalostí (činností). Pri absencii scenára by sme neboli schopní realizovať triviálne úlohy.

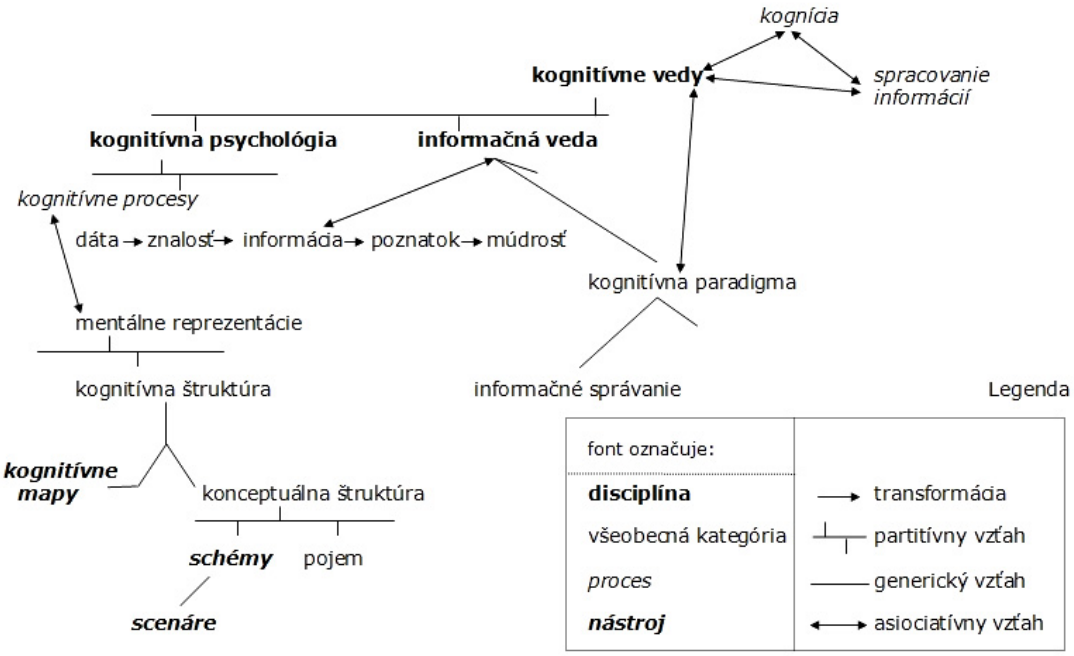
Analyzované pojmy kognitívnej paradigmy informačnej vedy a kognitívnych vied sú graficky reprezentované v konceptuálnom grafe (obr. 2). V predchádzajúcej práci [24] sme sa zamerali na umelú inteligenciu (najmä reprezentáciu poznania a inteligentné systémy). V tomto grafe sa snažíme upozorniť na vzťahy medzi jednotlivými typmi pojmov (disciplína, všeobecná kategória, proces, nástroj). Pomocou vzťahov sú vyjadrené prepojenia medzi kognitívnymi vedami, kognitívnou psychológiou, informačnou vedou a kognitívnou paradigmou.

3 Genéza kognitívnej paradigmy a kritické hlasy

Informačná veda je pomerne mladá disciplína. Vyvíja sa od 50. rokov minulého storočia. Kognitívne vedy našli svoje miesto v informačnej vede približne od konca 80. rokov 20. storočia. Prierez počiatočného štádia kognitívnej paradigmy načrtol Ingwersen [31]. Podľa neho bol De Mey medzi prvými odborníkmi, ktorí využili poznatky kognitívnych vied v informačnej vede. Predstavil „*kognitívne hľadisko*“ na multidisciplinárnom seminári v Ghente (Workshop on Cognitive Viewpoint, 1977). Nasledovali ho B. C. Brookes, Belkin a ďalší vývoj ovplyvnili Debons a Griffith. Predmet informačnej vedy

bol rozšírený o *poznanie, porozumenie a význam*.

V oblasti kognitívnej paradigmy informačnej vedy sú známi mnohí autori. Patria medzi nich Ingwersen [31,33], Dervinová [16,17], Allen [1,2], Kuhltauová [36,37], Saracevic [46], Spinková [49,50], atď. Na Slovensku v tejto oblasti vyniká Steinerová [52,53,54].



Obrázok 2. Kognitívne základy informačnej vedy - konceptuálny graf.

Kognitívne hľadisko sa nevyhlo ostrej kritike. Zásadné kritické argumenty sprehľadňuje Hjørland [29] a Limbergová [39]. Upozornili na individualizmus, ktorý nezohľadňuje sociálny kontext. Hjørland presadzoval názor, že omnoho plodnejšia by mala byť sociologická orientácia a potvrdil to predostretím alternatívy, konkrétne doménovej analýzy. Hjørlanda podporil Frohmann, ktorý vyjadril námietky voči mentalizmu. Medzi zástancov diskurzívnej analýzy patria Palermitiová, Polityová a Taljaová. Syntézou kritík a kognitívnej paradigmy je kognitívny pohľad v sociálnom rámci. Odráža sa to aj v novšom rámci *„holistického kognitívneho hľadiska“* [33].

Ellis [39] upozornil na problematickosť analógie človeka a stroja (čiže funkcionalizmu, ktorý bol kritizovaný aj na pôde filozofie mysle). Stroj nedisponuje napr. komplikovanými, mnohorozmernými kognitívnymi procesmi ako človek (a už vôbec nie afektívnymi). Ľudské myslenie, rozhodovanie, či riešenie problémov nemožno eliminovať na následnosť algoritmov a kód.

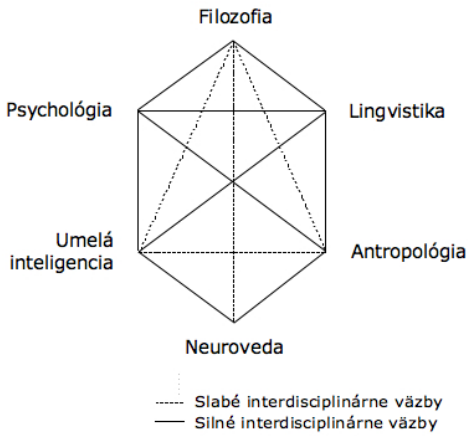
Otázne ešte ostáva, čo presne pomenúvajú abstraktné konštrukty mentalistického slovníka ako kognitívna štruktúra, kognitívne akty, kognitívne procesy a pod. Ako ich empiricky skúmať?

Kritické hlasy protivníkov kognitívnej paradigmy jej zatiaľ neuškodili, naopak, posilnili ju a inšpirovali. Prejavilo sa to aj v posune od individualizmu k vnímaniu človeka v sociálnom a kultúrnom kontexte, ale aj k holizmu.

4 Miesto informačnej vedy v kognitívnych vedách

Teoretické východiská informačnej vedy a kognitívnych vied majú spoločné základy. K vytvoreniu teórie informačnej vedy prispeli Bertalanffyho všeobecná teória systémov, Wienerova kybernetika, Shannonova teória informácií, Chomskeho teória syntaktických jazykových štruktúr, Newellova a Simonova teória spracovania informácií, atď.

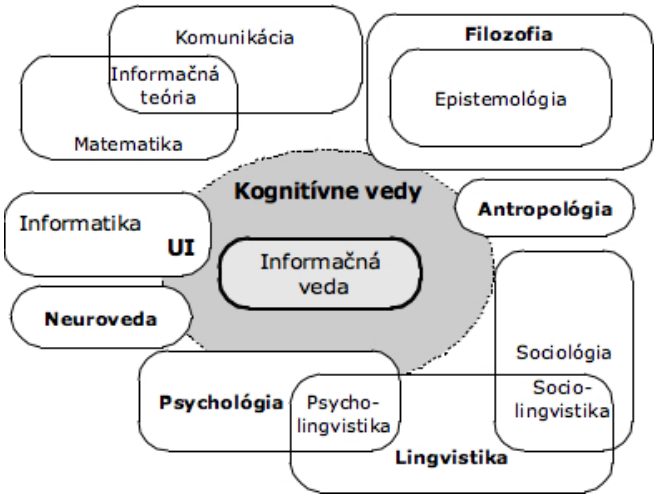
Vzťahy medzi jednotlivými disciplínami kognitívnych vied vizualizuje Gardnerov *kognitívny hexagram* (obr. 3).



Obrázok 3. Kognitívny hexagram znázorňuje silu interdisciplinárnych vzťahov [38].

V uvedenom hexagrame informačná veda absentuje. Následne predostrieme prístupy, ktoré informačnú vedu triedia ako súčasť kognitívnych vied alebo ju s nimi prepájajú.

Saracevic [38] navrhol model, v ktorom dal informačnú vedu do priameho vzťahu s informatikou, kognitívnymi vedami, komunikáciou a knihovníctvom. Ingwersen [31] vníma informačnú vedu ako súčasť kognitívnych vied. Na ilustráciu uvádzame model (obr. 4), ktorý znázorňuje disciplíny, ktoré ovplyvňujú informačnú vedu.



Obrázok 4. Disciplíny ovplyvňujúce informačnú vedu (prepracované a doplnené podľa Ingwersena [31]).

Model znázorňuje informačnú vedu ako disciplínu vnorenú do kognitívnych vied. Kognitívne vedy sú tvorené filozofiou, antropológiou, lingvistikou, psychológiou, neurovedou a umelou inteligenciou (v modeli sú zvýraznené tučným písmom) a ďalšími subdisciplínami. Informačná veda je ovplyvnená nielen kognitívnymi vedami, ale aj sociológiou, informatikou, matematikou, teóriou informácie a komunikáciou.

Súhlasíme s pohľadom, ktorý prezentuje informačnú vedu ako súčasť kognitívnych vied. V prieniku oboch disciplín sa nachádzajú kategórie spracovanie informácií, reprezentácia informácií, mentálne reprezentácie, kognícia, kognitívna štruktúra, konceptuálna štruktúra, kognitívne procesy, kognitívne štýly, informácia, poznatok, poznanie, význam, múdrosť a pod. Práve pri skúmaní a snahe o lepšie pochopenie týchto kategórií môžu nadviazať kognitívne vedy spoločný rozhovor s informačnou vedou.

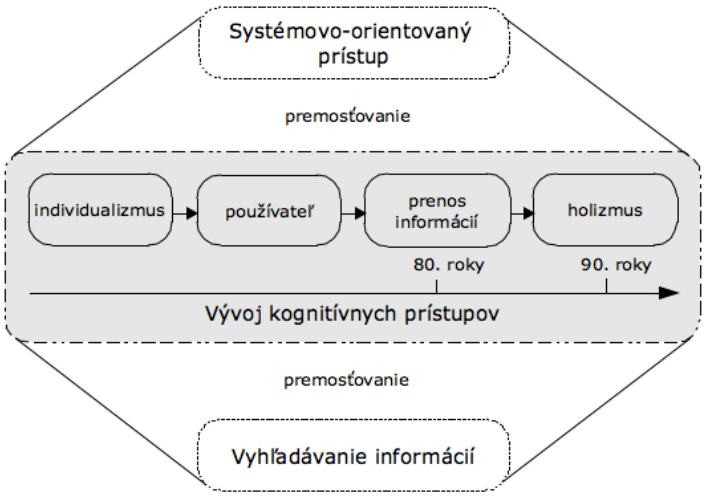
5 Kognitívna paradigma informačnej vedy

Kognitívne východiská informačnej vedy sa formovali dávno pred „boomom“ kognitívnych koncepcií. Tieto tendencie môžeme nájsť pri koreňoch informačnej vedy, ale aj na strane kognitívnych vied (najmä v kognitívnej psychológii, filozofii a umelej inteligencii).

Kognitívne prístupy v informačnej vede premosťujú systémovo-orientované prístupy s koncepciami vyhľadávania informácií [33]. Medzi základné východiská kognitívnej paradigmy patrí Taylorov model vytvárania informačnej potreby. Belkinova, Oddyho a Brooksova koncepcia ASK (Anomalous State of Knowledge, anomálny stav poznania) posunula systémovo-orientované hľadisko k orientácii na používateľa. Významné miesto mal aj princíp polyreprezentácie. K formovaniu kognitívnych základov informačnej vedy významne prispel Allen [1]. Rozlíšil kognitívne schopnosti, kognitívne procesy a kognitívne štýly. Jednotlivé kategórie podrobne analyzoval a poukázal na konkrétne možnosti ich aplikácie.

Vývoj kognitívnych prístupov znázorňuje obr. 5. Model je vytvorený na základe koncepcie, ktorú opísali Ingwersen a Järvelin [33]. Poukázali na zlomové body v genéze kognitívnej paradigmy, ku ktorým patril individualizmus a neskôr orientácia na používateľa. Ďalej sa v rámci kognitívnej paradigmy zdôrazňovali kvalitatívne aspekty prenosu a využívania informácií. Prenos informácií sa mal zdokonaľovať tvorbou modelov poznania používateľov, ktoré mali byť kompatibilné s konceptuálnymi reprezentáciami v informačných systémoch [1]. V súčasnosti sa záujem o procesuálne aspekty utlmil a prevláda skôr holistická orientácia. Holizmus poníma jednotlivca v kontexte a neoddeľuje ho od jeho okolia. Taktiež sa upustilo od myšlienky kompatibility, pretože kognitívne rozdiely (medzi ľuďmi, medzi ľuďmi a systémami, medzi systémami navzájom) sa zdajú byť neprekonateľné.

Kognitívne hľadiská sú pretavené do výskumu informačného správania (časť 5.1), interakčných prístupov (časť 5.2). Výskumy v týchto oblastiach sú aplikované na dizajn informačných systémov (časť 5.3).



Obrázok 5. Vývoj kognitívnych prístupov v informačnej vede.

5.1 Informačné správanie

Jedným z najvýraznejších smerov v informačnej vede je výskum informačného správania. Informačné správanie označuje interakciu človeka s informačným prostredím v sociálnom kontexte. Vzniká pri adaptácii človeka na prostredie, keď reaguje na konkrétne situácie pri riešení problémov. Informačné správanie možno vnímať aj ako „hľadanie významu“ [37], či „vytváranie zmyslu“ [16]. Batesová [4] vytvorila model, ktorý vysvetľuje, že informačné správanie je potrebné chápať ako sústavu vrstiev. Najnižšia vrstva je *chemická, fyzická, geologická a astronomická*; ďalej nasleduje *biologická* (genetika a etnológia); *antropologická* (fyzická a kultúrna); *sociálna a historická* (sociálne vedy); *kognitívna/konatívna/afektívna* (psychológia); *estetická* (umenie a literatúra); *spirituálna* (náboženstvo, filozofia, nachádzanie významu). Informačná veda sa v súčasnosti orientuje na sociálnu a historickú vrstvu, kognitívnu/konatívnu/afektívnu vrstvu a čiastočne aj chemickú, fyzickú a biologickú vrstvu. Rozšírenie a prehĺbenie poznatkov o ďalších vrstvách môže byť pre informačnú vedu veľmi užitočné a inšpirujúce.

Východiskom skúmania informačného správania je Dervinovej koncepcia *vytvárania zmyslu* („sense-making“) [16,17], ktorá vychádza z konštruktivizmu, situačného prístupu a holizmu. Vysvetľuje, ako človek v konkrétnej *situácii* (história, skúsenosť, horizont minulosti/prítomnosti) prostredníctvom *vytvárania zmyslu* premosťuje priepasť (*medzeru*) a získava *výstupy* v časovo-priestorovom rámci. Most je tvorený ideami, kogníciami, myšlienkami, postojmi, vierami, hodnotami, pocitmi, emóciami, intuíciami, spomienkami, príbehmi a rozprávaním. V medzere Dervinová rozpoznáva otázky, chaos, zmätok, hádanky a strach. V najnovšom ponímaní teória vytvárania zmyslu označuje všeobecnejšiu metodológiu a potrebu premosťovania odborov (interdisciplinárny prístup) [18].

Nahlová [40] predstavuje trichotomickú taxonómiu informačného správania *ACS* (Affective → Cognitive → Sensomotoric, Afektívny → Kognitívny → Senzomotorický). V interakcii človeka s informačným prostredím sa podľa ACS taxonómie prejavujú tri

zložky informačného správania: *afektívna, kognitívna a senzomotorická*. Nahlová kladie dôraz najmä na afektívnu zložku. V modeli informačného správania charakterizuje jednotlivé zložky na troch úrovniach: adaptácia, interakcia a personalizácia.

Veľmi inšpirujúce sú výskumy informačného správania v prostredí internetu. *CIBER* (Centre for Information Behaviour and the Evaluation of Research, Centrum pre informačné správanie a hodnotenie výskumu) [11], sa zameriava na výskumné projekty týkajúce sa rôznych používateľských skupín (žurnalisti, divadelníci, doktori), digitálnych knižníc, elektronických zdrojov a digitálneho zdravia. Informačnému správaniu v elektronickom prostredí sa venujú napr. Batesová [3], Nicholas et. al. [41,42] a Spinková a Jansen [51].

5.1.1 Modely informačného správania pri vyhľadávaní informácií

Modely vyhľadávania informácií vznikajú syntézou empirických alebo konceptuálnych analýz. Vzťah medzi informačným správaním, vyhľadávaním informácií a informačným správaním vymedzil Wilson [60]. Vo svojom modeli uvádza informačné správanie ako najvšeobecnejšiu kategóriu. Hierarchicky podradené je vyhľadávanie informácií a napokon najšpecifickejšiu oblasť predstavuje informačný prieskum (Ingwersen a Järvelin [33] na tomto mieste uvádzajú interaktívny informačný prieskum). Najznámejšie modely predstavili Ellis [19], Kuhltauová [36] a Wilson a Walsh [61].

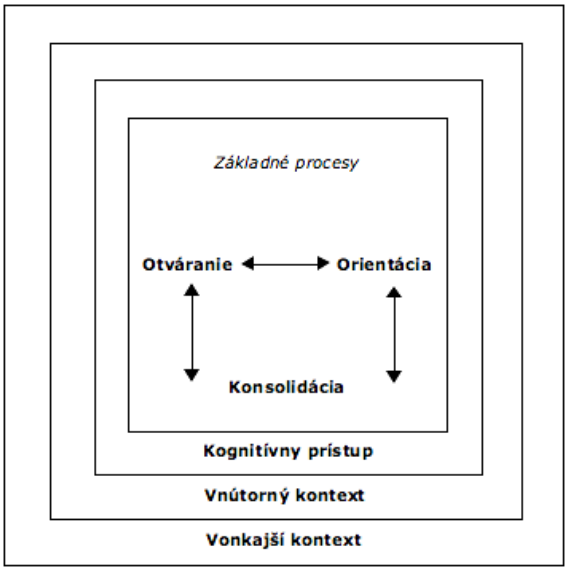
Ellis [19] v procesuálnom modeli pomenoval konkrétne charakteristiky: začatie, reťazenie, diferenciácia, extrakcia, overovanie a ukončenie. Na všeobecnejšej úrovni informačného správania určil prezeranie a monitorovanie. Kuhltauová [36] obohatila Ellisov model o *štádiovú* vrstvu (iniciácia, selekcia/prieskum, formulácia, zber, prezentácia) a *činnostnú* vrstvu (rozpoznanie, identifikácia/formulácia, zhromaždenie, doplnenie). Pôvodný model tak rozšírila o afektívne a kognitívne aspekty. Modely naznačujú niektoré zovšeobecnené procesy vyhľadávania informácií. Kladú čiastočne odpoveď na otázku, čo sa deje, keď človek vyhľadáva informácie, ale neberú do úvahy napr. situáciu (rozpoloženie) človeka, úlohu (riešenie problému), kontext a ďalšie interaktívne aspekty.

Wilson a Walsh [61] vychádzajú pri modelovaní informačného správania v procese vyhľadávania informácií z troch teórií: vyrovnávania sa so stresom, sociálneho učenia a rizika/úžitku. Tieto teórie podmieňujú aktivačné mechanizmy prítomné v informačnom správaní. Do procesu vchádzajú aj niektoré charakteristiky prostredia (zasahujúce premenné). Spätnoväzobný proces prepája informačné správanie pri vyhľadávaní informácií s osobou v kontexte a realizuje sa v ňom spracovanie a využívanie informácií. Wilson [60] sa neskôr pri modelovaní informačného správania inšpiroval procesmi riešenia problémov.

V súčasnosti sú veľmi obľúbené prístupy založené na (pracovných, každodenných) *úlohách*, záujmoch a *diskurzná analýza*.

Foster [22] predstavil *nelineárny model* (obr. 6), ktorý je založený na identifikácii procesov, kontextov a vzorcov správania a ich vzťahov. Proces vyhľadávania informácií je podľa neho nelineárny, dynamický, holistický a plynulý. Nový model zoskupuje činnosti do troch základných kategórií: *otváranie* (posun od orientácie k vyhľadávaniu), *orientácia* (identifikácia existujúceho a určenie smeru vyhľadávania) a *konsolidácia* (ladenie a poznanie momentu, kedy ukončiť prieskum). Fosterov model predstavuje aj interakčný aspekt informačného správania. Vymedzil tri *úrovne kontextuálnej interakcie*: vnútorný a vonkajší kontext a kognitívny prístup. Foster identifikoval štyri *kognitívne*

prístupy: flexibilný a adaptabilný, ktorý zdôrazňuje mentálnu agilnosť a vôľu adaptovať sa na rôzne kultúry; otvorený prístup bez predošlého rámca; nomadický prístup, v ktorom sa aktívne hľadajú rôzne spôsoby prístupu; holistický prístup zdôrazňujúci spájanie odlišných oblastí. Nelineárnosť modelu bola viacerými argumentmi spochybnená Wilsonom, ktorý vyvolal na túto tému diskusiu.



Obrázok 6. Fosterov nelineárny model informačného správania [22].

Líniu výskumov informačného správania charakterizuje rozmanitosť aspektov. Ich praktickú užitočnosť podmieňuje zohľadnenie situovanosti procesov a previazanosti na skúmané objekty. Subjektivita výskumu sa ukazuje ako problematická, ale nie neprekonateľná (napr. nástrojmi na validáciu). Medzi príklady aplikácie výsledkov patrí napr. výskum správania ľudí pri vyhľadávaní informácií v internete pri riešení každodenných problémov, ktorý bol využitý pri tvorbe informačných systémov pre komunity [27]. Skúmanie informačného správania pri kolaboratívnom prieskume v pracovnom kolektíve firmy bol zas využitý pri návrhu kolaboratívnych systémov [14]. Výsledky výskumov informačného správania sa môžu uplatniť pri tvorbe aplikácií, ktoré sú využívané človekom, kolektívom alebo komunitou pri riešení problémov a úloh.

5.2 Interakčný prístup

Interaktívny informačný prieskum sprehľadnili práce Beaulieuovej [5] a Savage-Knepshielda a Belkina [47]. K formovaniu tejto oblasti prispeli výskumy z prostredia bibliografických databáz. Neskôr sa ohnisko záujmu spolu s rozšírením internetu presunulo do priestoru webu (najmä digitálnych knižníc a prieskumových strojov). Prínosom pre interaktívny informačný prieskum bola Batesovej koncepcia „berrypicking“ (zbieranie lesných plodov) [3]. Na základe tejto koncepcie sa vystavali základy exploratívneho informačného prieskumu. Na výskumy koncového používateľa (interaktívne procesy a spätnú väzbu) sa zamerala Fidelová [21]. Ďalšie výskumy sa sústredili na vývoj a posun ohniska informačných potrieb, vnímanie, formulovanie a opis požiadavky.

Dôležitý prínos predstavujú výskumy kolaboratívneho informačného prieskumu a analýza interakcie založená na (pracovných) úlohách. V ďalšej časti poukážeme na základné modely interaktívneho informačného prieskumu.

5.2.1 Modely prieskumovej interakcie

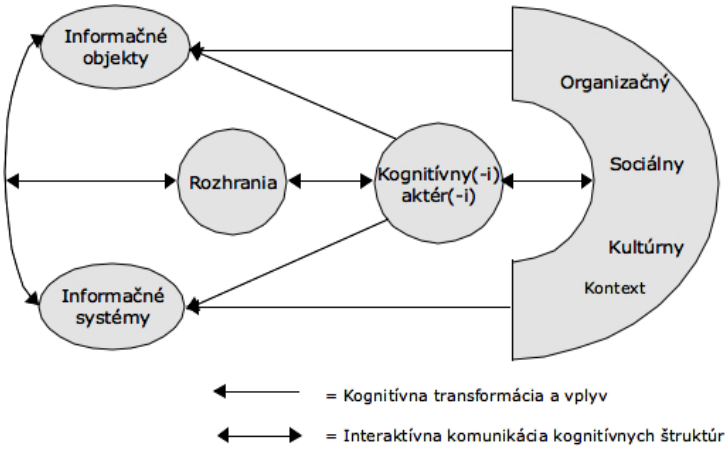
Interaktívny informačný prieskum je reprezentovaný *modelmi prieskumovej interakcie*. Ingwersen [31,33] prezentuje východiská informačnej prieskumovej interakcie v kontexte kognitívnej paradigmy. Prvé modely prieskumovej interakcie mali procesuálny charakter a vymedzovali konkrétne prípady vyhľadávania (za využitia boolovej logiky). Metodológia bola striktne ohraničená a exaktná z praktických dôvodov – online prieskum bol finančne náročný.

Medzi východiskové modely prieskumovej interakcie patrí Belkinov *epizodický model interakcií* pri vyhľadávaní informácií [6]. Základnými prvkami modelu sú používateľ, informačný zdroj a sprostredkovateľ. Naznačuje procesy reprezentácie, porovnávania, interakcie, hodnotenia a úpravy.

Saracevic [46] vytvoril *model stratifikovanej interakcie* pozostávajúci z jednotlivých úrovní a ich vrstiev. Na úrovni *používateľa* odlišuje kognitívnu, afektívnu a situačnú vrstvu. Na úrovni *počítača* sú vrstvy technického spracovania, spracovania (programy, algoritmy) a obsahu. Medzi používateľom a počítačom je rozhranie, smerom ku ktorému sa realizuje adaptácia oboch úrovní. Z hľadiska úrovní komunikácie odlíšil 3 vrstvy: *povrchovú vrstvu spracovania* (výpočtové spracovanie dát), *interaktívnu kognitívnu vrstvu* (vnímanie informácií počas interakcie človek-počítač) a *situačnú vrstvu* (využitie informácií na splnenie pracovnej úlohy v kontexte okolia).

Spinková [50] zahrnula do svojho modelu *interakciu ako spätnoväzobný cyklus* a Xienová [62] navrhla *model interaktívneho vyhľadávania*, ktoré vníma ako súhru osobnostných štruktúr používateľa, zdrojov a situácie. Jadro modelu je tvorené analýzou cieľov, plánov a interaktívnych zámerov, ktoré ústia do vyhľadávacej stratégie.

Medzi najaktuálnejšie patrí *všeobecný analytický model vyhľadávania a prieskumu informácií* [32,33]. V modeli (obr. 7) Ingwersen a Järvelin naznačujú toky kognitívnej transformácie a vplyvu i interaktívnej komunikácie kognitívnych štruktúr, ktoré sa realizujú medzi informačnými objektmi, informačnými systémami, rozhraniami, kognitívnymi aktérmi a organizačným, sociálnym a kultúrnym kontextom.



Obrázok 7. Všeobecný analytický model vyhľadávania a prieskumu informácií [32,33].

Tieto a mnohé ďalšie modely sú vynikajúcim nástrojom na lepšie pochopenie procesov, ktoré sa realizujú vo vzťahu človeka a technológií v širšom kontexte. Ich poznanie umožňuje nielen modelovať a simulovať ich, ale aj podporiť využiteľnosť technologických nástrojov. Najväčší progres zaznamenávajú prístupy orientované na človeka/používateľa, interakciu človek počítač/agent, interakčný dizajn, adaptívny dizajn, využiteľnosť, informačnú architektúra a pod. Medzi konkrétne aplikácie na vyhľadávanie informácií, ktoré podporujú interaktivitu patria napr. KartOO [35] a Ujiko [59].

5.2.2 Analýza a hodnotenie interakcie

Východiská na analýzu interakcie určuje koncepcia Ingwersena a Järvelina [33], ktorí určili rámce vyhľadávania informácií, dizajnu prieskumu a hodnotenia. Ich model zahŕňa kontexty, v rámci ktorých naznačujú kritériá na ich hodnotenie. Ide o kontexty *informačného prieskumu* (kritériá odozvy, presnosti, efektívnosti, kvality informácií/procesov), *vyhľadávania* (kritériá využiteľnosti, kvality informácií/procesov), *pracovnej úlohy* (kritériá kvality informačných a pracovných procesov/výsledkov) a *sociálno-organizačný a kultúrny kontext* (kritériá sociálno-kognitívnej relevancie, kvality výsledkov pracovných úloh). Ďalšie, novšie koncipované miery predstavujú efektívnosť systému v rôznych dimenziách počas informačnej prieskumovej interakcie, kvalita vyhľadaných informácií, kvalita prieskumového procesu (napr. čas), spokojnosť a druhy využívaných postupov a taktík.

Najprehľadnejšie charakterizuje rámec na hodnotenie interakcie človeka s informačným systémom Brajnik [9]. Zohľadňuje aspekt využiteľnosti systému na dvoch úrovniach: prezentačnej a navigačnej. V prezentačnej úrovni určil kritériá hodnotenia: schopnosť učiť sa, flexibilita a robustnosť; na navigačnej úrovni: pozorovateľnosť, prehľadnosť operácií, predvídateľnosť, dostupnosť, obnoviteľnosť, ucelenosť a personalizovateľnosť.

Kritériá hodnotenia interakcie majú vychádzať z cieľov a predmetu výskumu. V prostredí informačnej vedy sú veľmi frekventované výskumy zamerané na digitálne knižnice a nástroje na vyhľadávanie a organizáciu informácií. Komplexnejší pohľad poskytujú výskumy informačného správania orientované na používateľa. Hodnotenie systému vychádza z podpory interakcie. Funkčnosť a podpora systému sú často v rozpore s očakávaniami používateľa. Práve v tomto bode môžu analýzy a hodnotenia systémov prispieť k ich efektívnejšej využiteľnosti.

5.3 Aplikácia kognitívnych vied v dizajne informačných systémov

Línia výskumov orientovaných na aplikáciu kognitívnych aspektov v dizajne je reprezentovaná najaktuálnejšími systémami. Progresívne sa vyvíjajú oblasti digitálnych knižníc, inteligentných informačných agentov, systémov na odporúčanie informácií, interakcie človek-agent/počítač/informácie, adaptabilného informačného prieskumu, multimediálneho prieskumu, atď. S aplikáciami sú úzko spojené problémy relevancie, sémantického webu, ontológií, spracovania prirodzeného jazyka, kognitívneho modelovania, modelovania založenom na jazyku, strojového učenia a pod.

V informačnej vede majú systémy podporujúce inteligentný prieskum takmer tridsaťročnú tradíciu. Spomenieme napr. expertné systémy *Monstrat* [7] a *I3R* (Intelligent, Interactive Information Retrieval; Inteligentný, Interaktívny Informačný Prieskum) [15].

Inteligentný prieskum sa špecializuje na aplikáciu kognitívnych vied (najmä umelej inteligencie) v informačnom prieskume. V súčasnosti je ohniskom záujmu vyvíjať moderné systémy na inteligentný prieskum v *CIIR* (Center for Intelligent Information Retrieval, Centrum pre inteligentný informačný prieskum) [12].

Odborníci volajú po syntéze výskumov orientovaných na človeka a dizajnu aplikácií. Johnstone et al. [34] skĺbili systémové modelovanie s výskumom informačného správania. Upozorňujú na metodológiu „mäkkých" systémov (SSM, Soft System Methodology) z 90. rokov minulého storočia, ktorá berie do úvahy človeka a kontext. Navrhujú *vývojový systémový model*, ktorý zohľadňuje informačné správanie ako na úrovni človeka, tak informačných technológií. Jedným z cieľov je poskytnúť rámec na tvorbu a implementáciu softvérových aplikácií.

Rasmussenová [43] stojí za myšlienkou *kognitívneho systémového inžinierstva*, ktoré si kladie za cieľ vyrovnať sa s komplexitou prostredníctvom riadenia adaptácií. Zameriava sa na tvorbu inteligentných systémov, ktoré simulujú kognitívne procesy človeka a podporujú jeho interakciu s informáciami.

Fidelová a Pejtersenová [20] vychádzajú z Rasmussenovej koncepcie *kognitívnej analýzy práce*, ktorá vysvetľuje informačnú prieskumovú interakciu prostredníctvom pracovných aktivít ľudí. Dizajn systému je podmienený *porozumením práce ľudí, informačného správania, kontextu a príčin ich aktivít*. Autorky vymedzili dimenzie kognitívnej analýzy práce, ktoré zahŕňajú pracovné prostredie, pracovnú doménu, úlohy, organizácie, rozhodnutí, stratégií, zdrojov a hodnôt používateľa. Podľa Fidelovej a Pejtersenovej nie je možné zahrnúť všetky aspekty komplexnosti interakcie do dizajnu systému, preto systémy vyžadujú určitú mieru adaptácie zo strany používateľa.

6 Namiesto záveru

Spomínané východiská sú potrebné pri syntéze hľadísk orientovaných na človeka a systémy. Vytvárajú koncepcie a modely, ktoré pri tvorbe nových, či zefektívňovaní existujúcich systémov zohľadňujú človeka (ľudí) a kontext. Známe aplikácie sú najmä z oblasti inteligentného informačného prieskumu, interakcie človek-stroj, kognitívneho systémového inžinierstva a pod.

Spoluprácu medzi inžinierskymi a sociálno-vednými disciplínami môže komplikovať odlišnosť ich jazykov. Tento problém bol v informačnej vede zdôraznený už pred viac ako tridsiatimi rokmi. Hendry [28] upozorňuje na nedostatočnú aplikáciu informačného správania v dizajne informačných systémov. Hlavný problém vidí v odlišnosti prístupov vo výskume a nedostatok aplikácie výsledkov z jednej oblasti do druhej. Výskum informačného správania a dizajnu informačných systémov sa vyznačujú dualitou niektorých aspektov: pozorovanie/intervencia, presný opis/kreativita alebo inovácia, abstrakcia/artefakty, zameranie na človeka/technológiu, orientácia na problém /riešenie, orientácia na analýzu/služby, disciplinarita/interdisciplinarita. Napriek tejto diverzite verím, že kognitívna paradigma informačnej vedy má čo ponúknuť inžinierskym disciplínam, aby ich výstupy boli nielen navrhované, ale aj čo najefektívnejšie využiteľné ľuďmi.

Poďakovanie: Tento príspevok vznikol za podpory grantovej agentúry VEGA SR v rámci grantovej úlohy 1/2481/05 Využívanie informácií pri informačnom správaní vo vzdelávaní a vede.

Literatúra

[1] Allen, B. L.: Cognitive research in information science: Implications for design. In: Williams, M. E., ed. *ARIST*. vol. 26. Medford, NJ: Learned Information, 1991, p. 3-37.

[2] Allen, B. L., Kim, Kyung-Sun: Person and context in information seeking: interactions between cognitive and task variables. In: *The New Review of Information Behaviour Research: Studies of information seeking in context*. vol. 2. Cambridge: Taylor Graham, 2001, p. 1-16.

[3] Bates, M. J.: *The Design of Browsing and Berrypicking Techniques for the Online Search Interface* [online]. 1989. [cit. 2006-04-02]. Dostupné na internete: http://www.gseis.ucla.edu/faculty/bates/berrypicking.html.

[4] Bates, M. J.: *Toward an integrated model of information seeking and searching* [online]. The Fourth International Conf. on Information Needs, Seeking and Use in Different Contexts, Lisbon, Portugal, September 11-13, 2002. c2002-2003. [cit. 2006-04-02]. Dostupné na internete: http://www.gseis.ucla.edu/faculty/bates/articles/info_SeekSearch-i-030329.html.

[5] Beaulieu, M.: Interaction in information searching and retrieval. In: Ingwersen, P., Järvelin, K.: *The Turn: Integration of Information Seeking and retrieval in Context.* Springer, 2005. p. 218.

[6] Belkin, N. J., Cool, C., Stein, A., Thiel, U.: Cases, scripts and information seeking strategies: On the design of interactive information retrieval systems. *Expert Systems with Applications* **9** (1995) 379-395.

[7] Belkin, N. J., Brooks, H.: Using discourse analysis for the design of information retireval interaction mechanism. In: *ACM/SIGIR Conf. Proc.* New York: ACM, 1983, p. 31-47.

[8] Belkin, N. J.: Information concepts for information science. In: *JoD* **34** (1978) 55-85.

[9] Brajnik, G.: Information seeking as explorative learning. In: *Proceedings of Mira 99: Evaluating Interactive Information Retrieval* [online]. 1999 [cit. 2006-01-07]. Dostupné na internete: http://www.dimi.uniud.it/~giorgio/papers/mira99.ps.gz.

[10] Buckland, M.: *The landscape of Information Science: the ASIS at 62*. c1998. Dostupné na internete: http://www.sims.berkeley.edu/~buckland/asis62.html.

[11] CIBER. In: *Homepage* [online]. c2005 [cit. 2006-04-23]. Dostupné na internete: http://www.ucl.ac.uk/ciber/ciber.php.

[12] CIIR. In: *Homepage* [online]. c2006. [cit. 2006-04-24]. Dostupné na internete: http://ciir.cs.umass.edu/.

[13] Cognitive Science Society. In: *Homepage* [online]. c2005. [cit. 2006-04-24]. Dostupné na internete: http://www.cognitivesciencesociety.org/.

[14] Collaborative and Multimedia Systems. In: *Microsoft Research* [online]. c2006. [cit. 2006-05-02]. Dostupné na internete: http://research.microsoft.com/coet/.

[15] Croft, W. B., Thomson, R. H.: I3R: A new approach to the design of document retrieval systems. *JASIS* **38**(6) (1987) 389-404.

[16] Dervin, B.: *An overview of sense-making research: concepts, methods and results to date.* [online]. International Communications Association Annual Meeting. Dallas, Texas, May 1983. [cit. 2005-02-04]. Dostupné na internete: http://communication.sbs.ohio-state.edu/sense-making/art/artdervin83.html.
[17] Dervin, B.: From the mind's eye of the user: the sense-making Qualitative-Quantitative Methodology. In: Jack D. Glazier, Ronald R. Powell. *Qualitative Research in Information Management*. Englewood, Col.: Libraries Unlimited, 1992, s. 61-84.
[18] Dervin, B.: Human studies and user studies: a call for methodological inter-disciplinarity. In: *Information Research* [online]. 9(1) paper 166, 2003. [cit 2006-01-05]. Dostupné na internete: http://InformationR.net/ir/9-1/paper166.html.
[19] Ellis, D.: A behaviourial approach to information retrieval system design. In: *JoD* **45**(3) (1989) 171-212.
[20] Fidel, R. Pejtersen, A. M.: From information behaviour research to the design of information systems: the Cognitive Work Analysis framework. In: *Information Research* [online]. 10(1), October 2004 [cit 2006-01-05]. Dostupné na internete: http://InformationR.net/ir/10-1/paper210.html.
[21] Fidel, R.: Searchers‘ selection of search keys I-III. In: Ingwersen, P., Järvelin, K.: *The Turn: Integration of Information Seeking and retrieval in Context.* Springer, 2005. p. 218.
[22] Foster, A. E.: A non-linear model of information seeking behaviour. In: *Information Research* [online]. 10(2), January 2005 [cit 2006-01-05]. Dostupné na internete: http://InformationR.net/ir/10-2/paper222.html.
[23] Gál, E.: Filozofia mysle a kognitívne vedy. In: RYBÁR, Ján et al. *Kognitívne vedy.* Bratislava: Kalligram, 2002. s. 21-46.
[24] Grešková, M.: *Aplikácie kognitívnych vied v informačnom prieskume*. [Diplomová práca]. Univerzita Komenského v Bratislave. Filozofická fakulta; KKIV. Školiteľ: Doc. PhDr. Jela Steinerová, PhD. Bratislava: FFUK, 2005. 80 s.
[25] Grešková, M.: Cognitive foundations of information science. In: *Využívanie informácií v informačnej spoločnosti. Information Use in Information Society. Proc. of the international conf.* Ed. J. Steinerová, J. Šušol. Bratislava, Slovakia, October, 10-11 2006. Bratislava: Centrum VTI, 2006. 228 s. Dostupné na internete: http://old-kkiv.fphil.uniba.sk/viis/VIISzbornikkomplet.pdf.
[26] Grešková, M.: Kognitívna paradigma informačnej vedy. In: *Kognice a umělý život VI* (zost. Jozef Kelemen, Vladimír Kvasnička). Opava: Slezká univerzita v Opavě, 2006, s. 149-157. Dostupné na internete: ftp://math.chtf.stuba.sk/pub/vlado/CogSci_AL_Trest/Greskova.pdf.
[27] Help-seeking in an electronic world : how people use the Internet for everyday situations (1998-2000). In: *IBEC* [online]. Seattle, WA: University of Washigton, 2006. [cit. 2006-05-02]. Dostupné na internete: http:/ibec.ischool.washington.edu/default1024.aspx?subCat=Help-seeking&cat=Projects.
[28] Hendry, D. G.: *Lost in translation: towards mutual understanding.* ASIST' 2005: Sparking Synergies: Bringing Research and Practice Together, October 28 - November 2, 2005, Charlotte, North Carolina. Dostupné na internete: http://faculty.washington.edu/dhendry/docs/asist2005.pdf.

[29] Hjørland, B.: Cognitive views in Library and Information Science (LIS). In: *Core Concepts in Library and Information Science (LIS).* c2006. Last edited 1-11-2006. [cit. 2006-12-17]. Dostupné na internete: http://www.db.dk/bh/Core Concepts in LIS/articles a-z/cognitive_paradigm.htm.
[30] Hjørland, B.: Documents, memory institutions and information science. *JoD*, **56**(1) (2000) 27-41. [cit. 2006-03-29]. Dostupné na internete: http://www.db.dk/bh/publikationer/Filer/Documents_memory%20institutions%20and%20IS.pdf.
[31] Ingwersen, P.: *Information retrieval interaction.* [online]. London: Taylor Graham, 1992. 246 p. [cit. 2006-02-04]. Dostupné na internete: http://www.db.dk/pi/iri/.
[32] Ingwersen, P., Järvelin, K.: Information seeking research needs extension towards tasks and technology. In: *Information research* [online]. 10(1), 2004 [cit. 2006-01-07]. Dostupné na internete: http://informationr.net/ir/10-1/paper212.html.
[33] Ingwersen, P., Järvelin, K.: *The Turn: Integration of Information Seeking and retrieval in Context.* Springer 2005.
[34] Johnstone, D., Bonner, M., Tate, M.: Bringing human information behaviour into information systems research: an application of systems modelling. In: *Information Research*, 9(4) paper 191, 2004. Dostupné na internete: http://InformationR.net/ir/9-4/paper191.html.
[35] *Kartoo* [online]. c2006 [cit. 2006-05-02]. Dostupné na internete: http://www.kartoo.com/.
[36] Kuhltau, C. C.: Inside the Information Search Process: information seeking from the user's perspective. In: *JASIS* **42**(5) (1991) 361-371.
[37] Kuhltau, C. C.: *Seeking meaning: a process approach to library and information services.* Norwood, NJ: Ablex, 1993, 199 p.
[38] Lima, G. A. Borém, Raghavan, K. S.: Classification of the web: information retrieval and cognitive research. *Knowledge organization* **31** (2004) 98-105.
[39] Limberg, L.: Phenomenography: a relational approach to research on information needs, seeking and use. In: Lars Höglund, Tom Wilson (eds.). *The New Review of Information Behaviour Research.* Vol. 1. London: Taylor Graham Publishing 2000.
[40] Nahl, D.: A Conceptual Framework for Explaining Information Behavior. In: *Studies in Media & Information Literacy Education* [online]. 1(2), may 2001 [cit 2006-01-05]. Dostupné na internete: http://www.utpjournals.com/jour.ihtml?lp=simile/issue2/nahlfulltext.html.
[41] Nicholas, D., Huntington, P., Williams, P.: Digital information consumers: characteristics and information seeking behaviour. In: Spink, A. (ed.), *New Directions in Human Information Behavior*. Kluwer Academic Publishers, 2005.
[42] Nicholas, D., Huntington, P., Williams, P., Dobrowolski, T.: Re-appraising information seeking behaviour in a digital environment: bouncers, checkers, returnees and the like. In: *JoD* **60**(1) (2004) 24-39.
[43] Rasmussen, J.: Information processing and human-machine interaction: an approach to Cognitive Engineering. In: Fidel, R. Pejtersen, A. M.: From information behaviour research to the design of information systems : the Cognitive Work Analysis framework. In: *Information Research* [online]. 10(1), October 2004 [cit 2006-01-05]. Dostupné na internete: http://InformationR.net/ir/10-1/paper210.html.
[44] Ruisel, I.: *Inteligencia a myslenie*. Bratislava: Ikar, 2004. 432 s.

[45] Ruisel, I., Ruiselová, Z.: Vybrané problémy psychológie poznávania. In: Ruisel, I.: *Kognitívna psychológia.* 2001. [cit. 2006-04-25]. Dostupné na internete: http://math.chtf.stuba.sk/kog_vedy.htm.
[46] Saracevic, T.: The stratified model of information retrieval interaction: Extension and applications. In: Schwartz, C., Rorvig, M. ASIS'97: *Proceedings of the ASIS.* vol. 34. Silver Spring (Maryland): ASIS, 1997, p. 313-327. [cit. 2006-01-06]. Dostupné na internete: http://www.scils.rutgers.edu/~tefko/ProcASIS1997.doc.
[47] Savage-Knepshield, P., Belkin, N. J.: Interaction in information retrieval: trends over time. In: Ingwersen, P., Järvelin, K.: *The Turn: Integration of Information Seeking and retrieval in Context.* Springer 2005, p. 218.
[48] Soergel, D.: *An Information Science Manifesto: ASIS Award of Merit Acceptance Speech.* Washington, DC, November 5, 1997. [cit. 2006-03-31]. Dostupné na internete: http://www.dsoergel.com/cv/B64.pdf.
[49] Spink, A., Cole, Ch.: *New Directions in Cognitive Information Retrieval*. Berlin: Springer 2005.
[50] Spink, A.: Information science: a third feedback framework. *JASIS.* **48**(8) (1997) 728-740.
[51] Spink, A., Jansen, B. J.: *Web Search: Public Searching of the Web*. Springer 2004.
[52] Steinerová, J.: *Informačné správanie: pohľady informačnej vedy*. Bratislava: CVTI 2005.
[53] Steinerová, J.: Kognitívne a sociálne prístupy informačnej vedy. In: *Zborník Filozofickej fakulty Univerzity Komenského: Knižničná a informačná veda*. Ročník XIX. Bratislava: Univerzita Komenského, 2002, s. 111-128.
[54] Steinerová, J.: *Teória informačného prieskumu*. Bratislava: Slovenská technická knižnica – Centrum VTI SR 1996.
[55] Sternberg, R. J.: *Kognitivní psychologie.* Praha: Portál, 2002, s. 133-179.
[56] Stillings, N. A. et al.: *Cognitive Science: An Introduction* [online]. Cambridge, Mass.: MIT Press, 1995. [cit. 2004-11-30]. Dostupné na internete: http://www.bpl.org/electronic/netlibrary.htm.
[57] Šlapák, O.: Dáta, informácie, znalosti. In: *E-logos: electronic journal for philosophy* [online]. 2003. [cit. 2006-05-03]. Dostupné na internete: http://nb.vse.cz/kfil/elogos/miscellany/slapa103.pdf.
[58] Thagard, P.: *Stanford Encyclopedia of Philosophy* [online]. c2004. [cit. 2004-12-10]. Heslo Cognitive Science. Dostupné na internete: http://plato.stanford.edu/entries/cognitive-science/.
[59] *Ujiko* [online]. c2006 [cit. 2006-05-02]. Dostupné na internete: http://www.ujiko.com/v2a/flash.php?langue=en.
[60] Wilson, T.D.: Models in information behaviour research. In: *JoD*. **55**(3) (1999) 249-270. Dostupné na internete: http://informationr.net/tdw/publ/papers/1999JDoc.html.
[61] Wilson, T. D., Walsh, C.: Information behaviour: an interdisciplinary perspective. In: Wilson, T. D. Models in information behaviour research. *JoD* **55** (1999) 3. [cit. 2006-04-20]. Dostupné na internete: http://informationr.net/tdw/publ/papers/1999JDoc.html.
[62] Xie, H.: Shifts of interactive intentions and information-seeking strategies in interactive information retrieval. In: *JASIS* **51**(9) (2000) 841-857.

Racionalita z pohledu logiky

Petr Jirků[1]

Abstrakt. Zabýváme se problémem racionality, který nahlížíme především v kontextu logiky. Je charakterizován vývoj logického myšlení a vyzvednuty hlavní mezníky za poslední dvě milénia. Upozorňujeme na fakt, že logický pohled na racionální myšlenkové konstrukce byl po většinu doby statický a všímáme si snah porozumět dynamice racionální inference. Konečně upozorňujeme na metody zkoumání cest inference a dedukce, tak jak je provádějí lidé.

1 Úvod

Začneme etymologií slova racionalita. Termín *racionální* pocházející z latinského *ratio* znamená rozumný, rozumový, odůvodněný, poměrný, přiměřený. V matematice jsou čísla racionální taková čísla, která vyjadřují poměr dvou čísel celých.

Racionalismus v původním slova smyslu je pokusem o řešení otázky původu a podstaty lidského poznávání. Je protikladem sensualismu. Byl prapůvodně užíván pro označení směru v teologii, který odvozoval ideu Boha z rozumového poznání (deismus) nebo dokonce úplné spoléhání na rozum a na poznávání prostřednictvím rozumu. (Srovnej racionálně podkládané středověké důkazy Boží existence, např. Anselmův důkaz, který byl parafrázován Kurtem Gödelem v kontextu formální modální logiky [2].) Důraz se klade na samočinnost rozumu, apriornost poznatků a deduktivnost pravd. V širším smyslu jde o zdůrazňování významu vědy, vzdělání a intelektu.

K velkým racionalistům patřili Sókratés, Platón i Aristotelés, samozřejmě stoikové, scholastikové, René Descartes, jistě Gottfried Wilhelm Leibniz, John Locke a také Immanuel Kant, abychom jmenovali aspoň některé významnější klasické reprezentanty. V moderní době je třeba po vídeňském kruhu jmenovat Lakatose, Kuhna, Laudana a další.

V našem textu se budeme racionalitou zabývat především z pohledu logického a to jak pohledu formální logiky tak i logiky neformální.

2 Racionalita

Racionalita je schopnost konat a myslet cestami, které jsou vedeny rozumem. Je to mj. hledání důvodů a důsledků argumentů a myšlenkových závěrů a rozhodnutí. Důvod na obecné úrovni je intencionální fenomén. Racionalita je pak univerzální v tom smyslu, že není nic takového jako racionalita pro něco či pro někoho, a co jakožto hledání důvodů a důsledků, není racionalitou pro něco či pro někoho jiného. Paradoxem racionality je, že racionalita nemá důvod, nebo lépe, racionalita nepotřebuje důvod. Naše chování, rozhodování a usuzování buď je nebo není racionální. Otázka, zda je racionální usuzovat nebo jednat racionálně nemá z tohoto pohledu dobrý smysl. O konceptu racionality z hlediska chování, jazyka a logiky pojednáváme s J. Kelemenem v publikaci [1]. Koncept racionality se v dějinách měnil. V sókratovském pojetí byl důraz kladen především na koherentnost myšlenkových procesů, Galileovská racionalita pak chápala pravdu jako korespondenci (vědeckých) poznatků a stavů věcí. Pozdější vývoj zejména v minulém

[1] Univerzita Karlova, Katedra logiky FF, 116 42 Praha 1, Celetná ul. 20, E-mail: petr.jirku@ff.cuni.cz

století ústy Karl Poppera podrobil toto pojetí kritice. Dnes možná ani tak nejde o klasické odhalování "věčných" pravd skrytých ve věcech, ale spíš o vytváření (konstruování) nové skutečnosti vzhledem k danému diskurzu.

2.1 Logika, přirozený jazyk a úskalí komunikace

Logika se projevuje skrze jazyk, ale bývá často v jazyce spíše skryta. V běžném usuzování jde o přirozený jazyk. Přirozený jazyk ale, jak ukazuje zkušenost, není logický. Je proto pozoruhodné, že se (aspoň občas) vůbec dorozumíme. Příklady naznačí některá úskalí a jejich rozmanitost. Když například „chytrá horákyně" přelstí krále, fandíme jí, i když vlastně spíš než o chytrost horákyně jde o hloupost krále, že se posunem významu termínů nechal tak snadno obelstít. Když například politik jedním dechem redaktorce v televizním pořadu na otázku, zda ve svém postu mohl pro danou věc udělat víc, prohlásí, že jistě „*mohl, ale že si myslí, že udělal maximum"*, je hned zřejmé, že se onen dotazovaný chtěl spíš z otázky vykroutit než na ni korektně odpovědět. Neúplná odpověď nebo vágní a zamlžená odpověď nebývá vždy jen projevem neschopnosti vydat korektní formulaci. Často je to spíše záměr otupit hrot ostře formulované otázky. Zamlčování informací – logik by raději řekl faktů - (ať vědomé či nevědomé), eufemismy pro částečné negativní odpovědi či dokonce pro nepravdy, v tom jsou mnozí lidé mistry. To je ale záležitost, která přesahuje prostor vyhrazený (formálním) logickým zkoumáním, to je spíš prostor pro úvahy psychologické.

Naprostá většina lidí v běžných životních situacích ovšem usuzuje logicky, tak proč je třeba logiku studovat? Není to v nejlepším případě ztráta času? Je pravda, že většinu běžných situací dobře zvládáme. Přeci jen se ale čas od času setkáváme se situacemi, kdy dovednost a poučení o logice zmůže mnoho. Jak obtížné je někdy pro běžného člověka (neškoleného v elementární logice a bez vhodných dovedností) správně vylovit negaci nějakého tvrzení se dvěma nebo dokonce třemi kvantifikátory. Už třeba jen poučení o dualitě disjunkce a konjunkce ve vztahu k negaci v podobě De Morganových zákonů je cenným nástrojem pro kontrolu správnosti některých úsudků. Příkladů je samozřejmě víc. Spojka „a" není vždy konjunkce. Jednoduchý příklad ukáže, že velmi často jde spíše o časovou následnost. Složený výrok „Upadl a vstal" nepochybně říká něco jiného než výrok „Vstal a upadl". Tato spojka „a" rozhodně není komutativní. Ale i v situaci, kdy nejde o časovou následnost, nýbrž o současnost, nemusí být vždy zřejmé, co přesně autor výroku měl na mysli. Podél kanadských dálnic se například můžete občas setkat s varovným nápisem: *Don´t drink and drive.* Co přesně znamená tento příkaz chápeme (chceme-li být vstřícní k autorovi) na základě dřívějších znalostí a výrok interpretujeme v tomto kontextu. Jen těžko např. přijmeme interpretaci, která řidiči nedovolí napít se vody. A to už vůbec nemluvíme o čase, tj. o tom zda jde o pití za jízdy, před jízdou, krátce před jízdou, dlouho před jízdou nebo dokonce až po jízdě.

Proto se při studiu jazykových výrazů a argumentů zabýváme presuposicemi, konverzačními implikaturami a dalšími skrytými předpoklady porozumění a dorozumívání. V jiných životních situacích to ale může být problematičtější. Opět příklad: Současný český vysokoškolský zákon např. v paragrafu 74 (1) jako podmínku pro profesorské a habilitační řízení vyžaduje, aby byla prokázána „pedagogická a vědecká nebo umělecká kvalifikace uchazeče". Fakt, že v přirozeném jazyce nepoužíváme závorky, má za následek, že se nabízejí aspoň dvě interpretace: $(p \wedge v) \vee u$ anebo $p \wedge (v \vee u)$. S tím se samozřejmě snadno vyrovnáme. Horší ovšem je, když např. vysoká škola v údajné -

možná i dobře míněné - snaze posoudit schopnost správně logicky usuzovat, klade uchazeči o studium otázky, které nejsou z logického hlediska v pořádku. Když zadavatel úlohy sám chybně hodnotí vztah vyplývání, je to zlé. A až hrozivé je, když v návodu jak vypracovat odpověď se studentovi doporučuje, aby odložil vše, co se případně v logice naučil. No, a když dokonce soud (byť to bylo v banální při) vydá rozsudek, který je kontradikcí, je to zlé. Jeden jihočeský soud nedávno vydal ve sporu rozvedených manželů, kteří se nedokázali dohodnout o tom jak se budou nadále stýkat se svými společnými dětmi toto rozhodnutí: *„Každý z obou rodičů převezme obě nezletilé děti v místě bydliště druhého rodiče a tamtéž druhému rodiči předá.“* Při znalosti faktu, že rodiče už spolu nemají stejné bydliště, to bude pro oba rodiče svízelné.

To jsou příklady situací, kdy se logika dostává do kontaktu s přirozeným jazykem a díky jeho vágnosti, neurčitosti a víceznačnosti nám vznikají problémy, jak výrazy jazyka správně interpretovat a jak potom správně odvozovat důsledky. Není to ale vada přirozeného jazyka. Zmíněné příklady spíš ukazují na neschopnost svoje myšlenky řádně formulovat s náležitou přesností. Nalézat odpovídající míru přesnosti je ovšem spíše záležitost cviku a schopnosti systematicky uvažovat a domýšlet důsledky. Tyto schopnosti je možné vždy zlepšovat. I přirozený jazyk takovou možnost poskytuje.

2.2 O co nám vlastně z hlediska logiky šlo, jde a půjde

Chápeme-li logiku jako *organon*, tj., jako nástroj usuzování, je třeba zkoumat, jaké má ten nástroj vlastnosti a k čemu je užíván, tedy co vlastně rozumíme usuzováním. Aristoteles v prvních analytikách říká:

Nejprve je třeba pojednat o tom, čeho se zkoumání týká a čí je to úkol; týká se důkazu a je to úkol dokazovací vědy. Nato je třeba určit, co je premisa a co termín a co sylogismus, který sylogismus je dokonalý a který nedokonalý.

První Aristotelova věta zde vystihuje přesně jeden z významných úkolů moderní logiky: porozumění tomu, co je to důkaz. Druhá věta se z dnešního pohledu týká partikulárního úkolu. Je tu ale potíž: Jsou různé náhledy. Populárním vymezením logiky ve starších učebnicích bývá vymezení logiky jako studia *metod správného myšlení* (Srovnej např. Krejčí, Dratvová, Pelikán a další). To ale není vyhovující. Myšlením se totiž zabývá řada disciplín (od psychologie, po neurofysiologii přes lingvistiku, kulturní antropologii, kognitivní vědu a další disciplíny). Je tedy třeba začít tím, že vymezíme *základní otázky*, které si v logice klademe. Jsou to, mezi jinými, jistě tyto otázky:

Co je to pravda? To není otázka po tom, které výroky jsou pravdivé, ale otázka po konceptu pravdy. Zde se logika setkává s filosofií. Pravdivost nebo nepravdivost vyjadřujeme v jazyce, proto zkoumáním vyjadřovacích prostředků musíme začít. Je třeba porovnat náhledy na pravdu (koherenční teorie, korespondenční teorie, ...)

Co z daných pravdivých (akceptovaných) výroků vyplývá a co z nich můžeme, popř. musíme, odvodit? Jde mj. o rozlišení nutnosti a kontingence. Přístup může být dvojí, a tak chceme porozumět tomu, co je to důkaz a dokazatelnost (syntaktický pohled) a co je to model či pravdivá interpretace (sémantický pohled). Co je vynucené a co nikoli. Co je možné a co nikoli. Ze středověké scholastiky jsme prostřednictvím Leibnize převzali koncept možných světů. V moderní vědě je to mj. diskuse o *raison d´être* myšlenkových pokusů (fyzika, kosmologie).

Jak souvisí dokazatelnost a pravdivost? Kurt Gödel nás přesvědčil, že koncept pravdy je bohatší než koncept důkazu, což silně ovlivnilo nejen moderní logiku samu, ale v posledním půlstoletí i další disciplíny jako např. lingvistiku a informatiku.

Jak obtížné je ty důsledky odvodit? To je otázka složitosti. Složitosti kalkulací. Zdá se, že v současné době je to otázka z nejaktuálnějších.

Jaké typy nededuktivních úsudků jsou ještě racionální? Dosud byl v logice kladen důraz na hledání věčných pravd (B. Pascal); jak je to s pravdami temporálními? Proč máme tolik logických kalkulů a jak se v nich orientovat? Co je v tomto kontextu racionalita? (Srovnej axiomy racionality podle P. Gärdenforse v dalším textu.)

K nejvýznamnějším konceptům, kterým se v logice snažíme porozumět, proto jistě patří *formální kalkul* a s tím spojené koncepty jako konzistence, nezávislost a další; *paradox* (self-reference, kontradikce, zapomenutá fakta, bezobsažnost nebo posunutí významu), asserce vs. negace (explozívnost negace, negace jako neúspěch, existence); *modalita a existence* možnost, nutnost a to, co není možné (*possibile* vs. *impossibile*); *algoritmus* (kalkul, výpočet, efektivnost, složitost) např. problém čtyř barev a velká Fermatova věta a zajisté *nekonečno* (základy matematiky).

O každém z uvedených témat by bylo možné konat více než semestrální přednášku. Pro náš účel alespoň rekapitulujeme hlavní mezníky v dějinách logiky a jejího okolí, neboť logika má, podobně jako matematika, tu pozoruhodnou vlastnost, že je ortogonální vůči ostatním disciplínám, prostupuje je. K těm patrně nejvýznamnějším mezníkům reprezentovaným význačnými velikány patří:

Tháles Milétský	racionální ospravedlnění
Pythágorás	čísla (*arithmos*) jako základ
Aristotelés	formální logika, *sylogismus, subjekt a predikát, modality*
Eukleidés	axiomatická metoda
Al Chvárizmí	algoritmus (postupy, posloupnost akcí)
Descartes	systematické pochybování
Newton	kalkul, (malé) nekonečno
Leibniz	možné světy, myšlení jako kalkul
Boole	booleovská logika
Bolzano	nekonečno, množství (množina)
Frege	predikátová logika
Hilbert	úsilí o prevenci paradoxů (axiomatická metoda v matematice)
Russell	teorie typů, Principia Mathematica
Gödel	pravdivost vs. dokazatelnost, nedokazatelnost bezespornosti *koncept pravdivosti je bohatší než koncept dokazatelnosti*
Turing	teoretický koncept počítače, umělá inteligence
Church	lambda kalkul, vyčíslitelnost *(computability)*
Chaitin	nový koncept náhody, algoritmická nestlačitelnost, algoritmická teorie informace

Ten výčet není samozřejmě úplný a je dokonce jistě i nespravedlivý, neboť ke každému uvedenému jménu lze přiřadit mnohá další jména jednotlivců, kteří ta velká témata utvářeli, ale to by se nám naše půlhodinová přednáška protáhla asi na dvě tisíciletí.

A co nás z tohoto pohledu čeká v právě započatém století? Patrně se budeme mj. muset zabývat též tématy jako logická epistemologie (dynamika znalostí, především potřeba vzít do úvahy čas). Která odvození lze ještě považovat za *racionální*? Co vlastně v tomto kontextu rozumíme racionalitou?

Dalším tématem, kterým bude třeba se v kontextu racionálního chování a rozhodování zabývat, bude opět logika jako nástroj "inženýrské" manipulace s poznatky a s tím související staronové otázky *Co je to stroj* a *Je člověk stroj?* V moderním hávu je to koncept umělé inteligence. Je totiž čas zhodnotit koncept Allana Turinga a s tím související hlubší porozumění složitosti, Chaitinův [11] koncept náhody jako algoritmické nestlačitelnosti a z toho vyplývající závěr, že (matematická) pravda může vykazovat charakteristiky náhody, nestrukturovanosti a nesrozumitelnosti. Z prací J. Matiasieviče a J. Jonese na 10. Hilbertově problému navíc vyplývá, že se to týká dokonce i elementární aritmetiky, jmenovitě aritmetiky v podání Guiseppe Peana. To jsme ostatně zažili i v případě Gödelově.

Budeme si také klást otázku, jak povstává logika a konceptuální myšlení z úrovně impulsů předávaných v síti neuronů. Jaké emergentní jevy v umělých neuronových sítích můžeme očekávat. A znovu nám půjde o hlubší porozumění nekonečnu, a to jak z pohledu zevnitř ven (např. velké kardinály), ale budeme si klást též otázku zda je možné nekonečno „vidět“ ve složitosti kalkulací? K tomu, že takové otázky si budeme dokonce klást naléhavě, nás už teď může vést skutečnost, že kolem nás roste síť strojů podřízených sice deterministickým pravidlům, která je ale topologickou strukturou složitostí v ní probíhajících procesů blízká architektuře lidského mozku. Další neméně vzrušující skutečností v souvislosti se starou Leibnizovou otázkou, zda my lidé jsme stroje či nikoli, je fakt, že jsme rozluštili lidský genom, který vlastně můžeme chápat jako složitou strukturu deterministických pravidel. Není bez zajímavosti, že počtem tři a půl miliardy genů jsme jen asi o jeden řád vzdáleni od počtu neuronů v průměrném lidském mozku.

Potřeba logiky je značná, její výskyt je nepatrný. Galileo Galilei v této souvislosti kdysi řekl: „*Nemůžeme vás naučit nic, můžeme vám jen vytvářet prostor pro to, abyste se naučili sami*.“ Logika je pro nás pevně spojena s matematickým náhledem na svět, ale není to nutně matematika, je to vyjevování struktury a souvislostí. Je to naše vidění světa a způsob našeho mluvení o světě.

2.3 Přirozený svět a svět překvapení.

Překvapení jsou zdrojem poznání. Odkud ale vyvěrají ona překvapení? Ukážeme to na několika příkladech, které nejsou klasickými paradoxy, tj. nejsou kontradikcemi ve smyslu logiky, ale naznačují jakýsi druh konfliktu rozumu se zjištěnou skutečností:

Již dávno jsme např. dobře poznali, že není možné rozdělit úhel na tři stejné díly. Když se to poprvé v životě dozvíme, tak nás to možná zaujme i třeba překvapí, ale ne příliš, protože si uvědomíme, úloha je umělá v tom smyslu, že se jedná o *omezené prostředky* a kdybychom měli k dispozici další pomůcky, například papír a nůžky, neměli bychom větší problémy. Skutečným překvapením, možná až velkým šokem pro starověké myslitele byl objev, že strana a úhlopříčka čtverce jsou nesouměřitelné. Prostor se ukázal být *podivným*. K tomu nás dovedl rozum Zcela jiným druhem překvapení je to, co známe pod názvem vězňovo dilema. To je pozoruhodné, ale ne tak dramatické protože jde o *rozhodování při neúplné informaci*. Nemožnost perfektní demokratické volby (*Arrowův teorém*) vlastně akceptujeme, protože to je zjištění, že standardní metodou hlasování

vlastně hodnotíme něco jiného než jsme původně chtěli. Problém obchodního cestujícího nás upozorňuje na *dramaticky rostoucí složitost výpočtu*, perpetuum mobile na *skrytý princip zajišťující soudržnost či konzistenci teorie*. To už je z logického hlediska zajímavější. Nemožnost důkazu konzistence aritmetiky, problém zastavení (*Halting problem*) v teorii algoritmů, relace neurčitosti kvantové fyziky (*není možné současné přesné měření některých veličin*) a myšlenkové pokusy obecně (*Maxwellův démon, Schrödingerova kočka, Einsteinův výtah, Newtonova kytice, ...*) to už jsou skutečné výzvy pro rozum a racionalitu. Je to jen kratičký a hlavně nesystematický namátkový výběr překvapení, která jsme zažili. Je to výčet pestrý, jistě se dočkáme překvapení dalších. Každý z příkladů naznačuje konflikt jiného druhu. Jde o situace, kdy rozum je překvapen. Překvapení se od sebe ale významně liší a to nejen tématem, ale i formou a především silou překvapení. Některá z takových překvapení se dotýkají základních logických konceptů a principů, zejména pak možnosti a nutnosti. Co je možné a co nikoli, co je to nutnost? V logice např. často užíváme princip *reductio ad absurdum, ale co je vlastně absurdní?*

Překvapení, které nám v logice v minulém století přichystal Kurt Gödel patřilo k největším, proto si zjednodušenou formou připomeneme oč šlo.

2.3.1 Gödelovská úloha

Kdysi překvapivý výsledek logických zkoumání Kurta Gödela, známý jako *věta o neúplnosti*, podle kterého dostatečně bohaté teorie nemohou být úplné - tj. musí v nich existovat pravdivá tvrzení, která nelze prostředky dané teorie dokázat - je často označován za *hlavní vnitřní hranici vědy*. V tomto příspěvku nemáme možnost detailně reprodukovat ideje, na nichž je Gödelův teorém založen, a které vynikající logik formuloval roku 1931, ale my se zde pokusíme čtenáři naznačit alespoň jejich odlesk prostřednictvím jednoduché a dobře srozumitelné úlohy formulované současným věhlasným logikem Raymundem Smullyanem. Úloha bude dnešnímu čtenáři jistě dobře srozumitelná. Budeme uvažovat jednoduchý výpočtový stroj (počítač), který umí mj. tisknout výrazy sestavované z těchto pěti znaků: ~ P N ( ). Výrazem budeme samozřejmě rozumět libovolný konečný řetězec znaků. Dále budeme předpokládat, že mezi našimi vyjadřovacími prostředky máme intuitivně srozumitelný unární predikát *tisknutelný*(X), který je splněn, právě když výraz X je tisknutelný, tj. když výraz X má tu vlastnost, že bude strojem dříve nebo později vytištěn. Dále ještě definujeme *normu* výrazu. Normou výrazu X budeme rozumět výraz, který vznikne z výrazu X podle definice norma(X) = X(X); norma výrazu je tedy jakýmsi „zdvojením“ původního výrazu. Takže např. normou výrazu ~NNP~ je výraz ~NNP~(~NNP~). A konečně některé z výrazů, které (jak dále uvidíme) nás budou velmi zajímat, budeme nazývat *sentence*. Sentencí budeme rozumět jakýkoli výraz některého z následujících tvarů: P(X), PN(X), ~P(X), ~PN(X), kde X je libovolný výraz. Tímto už jsme připraveni uvést definici pravdivosti (pravdivé sentence) pro náš jednoduchý jazyk. Bude záviset na formě, tedy tvaru sentence. Řekneme, že sentence tvaru

P(X) je pravdivá, když výraz X je tisknutelný,
PN(X) je pravdivá, když norma výrazu X je tisknutelná,
~P(X) je pravdivá, když výraz X není tisknutelný,
~PN(X) je pravdivá, když norma výrazu X není tisknutelná.

Než zformulujeme gödelovskou otázku, stanovíme důležitý předpoklad, který říká, že náš stroj je dokonale *korektní*, což znamená, že všechny sentence, které vytiskne, budou pravdivé. To mj. znamená, že když stroj např. tiskne sentenci tvaru P(X), tak X je skutečně tisknutelná (může být strojem dříve nebo později vytištěna). Tento požadavek se samozřejmě týká pouze sentencí, stroj totiž může případně tisknout i nesmysly, tj. výrazy, které nejsou sentencemi, ale to nám nečiní žádnou obtíž. A co když nyní X je výraz, který je tisknutelný. Znamená to, že také P(X) je tisknutelná sentence? Ne nutně. Jestliže totiž X je tisknutelný výraz, pak P(X) je jistě pravdivá sentence, ale nikde nemáme zaručeno, že stroj může (je schopen) vytisknout všechny pravdivé sentence. Víme jen, že stroj nikdy nevytiskne nepravdivou sentenci (díky korektnosti). Takže teď již můžeme formulovat otázku:

Je principiálně možné, aby stroj vytisknul všechny pravdivé sentence?

Čtenář asi tuší, že odpověď je negativní. Abychom to ověřili, stačí nalézt pravdivou sentenci, kterou stroj nemůže vytisknout. To se nám v našem případě podaří např. tak, že nalezneme sentenci, která tvrdí svoji vlastní netisknutelnost, což je sentence, která je pravdivá právě tehdy, když není strojem tisknutelná. Takovým řešením je sentence

~PN(~PN)

neboť podle definice pravdivosti je tato sentence pravdivá právě tehdy, když norma výrazu ~PN není tisknutelná. Ale normou výrazu ~PN je ona sentence samotná. Tudíž naše sentence je pravdivá právě tehdy, když není tisknutelná. To znamená, že tato sentence je buď pravdivá a netisknutelná anebo je tisknutelná a nepravdivá. Druhá alternativa ovšem porušuje předpoklad korektnosti (stroj nikdy netiskne sentence, které nejsou pravdivé), tudíž naše sentence musí být pravdivá, ale stroj ji nikdy nemůže vytisknout.

Všimněme si ještě, že sentence PN(~PN) je nepravdivá (protože její negace je pravdivá). Avšak tato sentence je v našem systému (opět díky korektnosti) rovněž nedokazatelná. A tak sentence PN(~PN) je příkladem sentence, která je v daném systému *nerozhodnutelná*.

2.3.2 Varianta Gödelovy úlohy

Ideu metody gödelizace můžeme velmi zhruba naznačit tak, že mírně pozměníme úlohu diskutovanou výše. Čtenář, který zná roli přirozených čísel v gödelovském kódování může tento odstavec bez újmy vynechat.

Jazyk bude opět sestávat z pěti symbolů, ale tentokrát budeme místo levé závorky používat symbol 1 a místo pravé závorky používat symbol 0. Dále budeme v této variantě ještě pracovat s přirozenými čísly. Ta budeme reprezentovat dvojkovým zápisem (jako řetězce nul a jedniček) a pro naše účely ztotožníme přirozená čísla s numerály, které je reprezentují. Důležitou roli teď bude hrát pojem Gödelova čísla výrazu. Gödelovo číslo *g*(X) výrazu X definujeme následovně: Pěti jednotlivým znakům ~, P, N, 1, 0, postupně přiřadíme Gödelova čísla 10, 100 1000, 10000, 100000. Gödelovo číslo složeného výrazu pak získáme složením (konkatenací) odpovídajících Gödelových čísel znaků, z nichž se skládá. Takže např. výraz PNP má Gödelovo číslo 1001000100. Norma výrazu bude definována rovností *norma*(X) = X*g*(X). Například normou výrazu PNP je výraz

PNP1001000100. Sentence mají nyní stejný tvar jako v prvním případě, ale zapsaný v dvojkové notaci. Definice pravdivosti bude mírně pozměněna takto:

- PX je pravdivá, když X je Gödelovo číslo pravdivého výrazu,
- PNX je pravdivá, když X je Gödelovo číslo výrazu, jehož norma je tisknutelná,
- ~PX je pravdivá, když PX není pravdivá (X není Gödelovo číslo tisknutelného výrazu),
- ~PNX je pravdivá, když PNX není pravdivá.

Formulace úlohy se nemění: Nalezněte pravdivou sentenci, kterou stroj nemůže vytisknout. Teď je už také zřejmé, že řešení takto modifikované úlohy je výraz ~PN101001000.

Uvažujme teď poněkud komplikovanější jazyk elementární aritmetiky, tj. teorii prvního řádu, která je zamýšlena jako teorie přirozených čísel, tak jak jsou přirozená čísla v běžném platónském chápání matematiky pojímána. Můžeme postupovat analogicky, pro samu aritmetiku bude ale gödelizace podstatně složitější (využije se prvočíselných rozkladů), ale idea z uvedeného příkladu prosvítá. Gödelovy výsledky silně ovlivnily nejen naše porozumění logice a matematice, ale i mnohé další vědní disciplíny mezi něž zcela jistě patří i lingvistika a třeba také kognitivní věda, která se snaží porozumět lidské mysli. Snad nám Smullyanova úloha pomohla osvětlit Gödelovu genialitu.

Naskýtá se teď přirozená otázka: Je to obecná vlastnost anebo náhodná vlastnost právě tohoto jednoduchého jazykového systému s právě touto definicí pravdivosti? Dnes dobře víme, že nikoli. Díky Kurtu Gödelovi a jeho rafinované metodě gödelizace (kódování formulí přirozenými čísly - viz dodatek) víme, že je to naopak principiální vlastnost dostatečně bohatých formálních systémů. Máme přitom bohatosti systému rozumět tak, že jde o systémy, které jsou složité? Co ale zde složitost znamená? Logikové vědí, že bohatost zde znamená, že uvažovaná teorie obsahuje (dá se v ní vyjádřit) elementární aritmetika neboli naše platónská představa o přirozených číslech.

K typickým vlastnostem klasických logických systémů patří mj. to, že dobře utvořené formule mají konečnou délku (ale může jich být nekonečně mnoho); pravidla odvozování mají jen konečně mnoho premis; jsou jen dvě pravdivostní hodnoty; všechny výrokové operátory (spojky) jsou extenzionální (truth-functional); platí princip *tertium non datur,* tj. formule $(p \vee \neg p)$ je dokazatelná, i když ani (p) ani $(\neg p)$ nejsou samostatně dokazatelné. Kontradikce je explozivní, plyne z ní cokoli; nebo jinak: obsahují-li axiomy spor, tak cokoli je teorémem; dokazatelnost je monotónní; predikáty mají za argumenty pouze individua (nikoli další predikáty); obor individuí je (*ex definitione*) neprázdný, neboli v každé interpretaci existuje aspoň jedno individuum; uvnitř predikátů není žádná struktura (pouze uspořádání argumentů); množina dobře utvořených formulí (množina symbolů a gramatických pravidel) je rozhodnutelná; což se týká i důkazů; standardní pravdivostní funkce jsou totální; predikáty jsou totální funkce z oboru individuí do pravdivostních hodnot; (formální) jazyk je uzavřený na gramatiku i na pravidla odvozování; formule, které jsou teorémy, jsou typu ano/ne (srovnej s fuzzy logikou).

2.4 Dedukce, indukce, abdukce

Pravidla deduktivní a induktivní inference jsou dobře známá, pravidla abduktivní již méně. Příklady:

a) Všichni králíci v tomto klobouku jsou bílí.
   Tito králíci jsou z tohoto klobouku.
   Tudíž: Tito králíci jsou bílí.
b) Tito králíci jsou z tohoto klobouku.
   Tito králíci jsou bílí.
   Tudíž: Všichni králíci v tomto klobouku jsou bílí.
c) Všichni králíci v tomto klobouku jsou bílí.
   Tito králíci jsou bílí.
   Tudíž: (*To je proto, že*) Tito králíci jsou z tohoto klobouku.

Příklad a) je typickým příkladem deduktivního úsudku, který garantuje pravdivost za předpokladu pravdivosti premis. Příklad b) je typickým příkladem induktivního úsudku, kde pravdivost již nemusí být garantována při pravdivosti premis. Příklad c) je pak příkladem abduktivního úsudku.

Abdukce je velmi účinné rozšíření inference. Termín abduktivní usuzování je velmi starý (srovnej Aristoteles), ale systematicky byl studován až ve dvacátém století Charlesem Sandersem Peircem a dalšími. V aristotelské logice je abdukce (apagogé) chápána jako sylogismus jehož hlavní premisa je jistá, ale vedlejší premisa je jen pravděpodobná. Podle Peirce je to druh inference, který přináší explanatorní hypotézy spíš než výsledek deduktivního či induktivního pravidla. Abduktivní usuzování je v tomto smyslu [8] usuzování z neúplných znalostí nebo hypotetické usuzování. V protikladu k dedukci se abdukce týká plauzibilnosti závěrů, nikoli jejich platnosti.

Kritéria síly abduktivních závěrů jsou především tato: Jak dobrá je hypotéza *H* sama, nezávisle na alternativách. Jakou decizivní přednost má hypotéza *H* před svými alternativami. A konečně jak důkladné či úplné bylo prohledávání prostoru alternativ. Z pragmatického hlediska je pak důležitá cena, kterou platíme za chybný závěr proti správnému a jaká je potřeba učinit vůbec nějaký závěr jmenovitě ve světle možnosti získat novou evidenci pro další rozhodování.

Formálněji můžeme abdukci popsat následovně: Když *B* je informace či znalost, která je pevným zázemím úsudku (*background knowledge*) a *G* je množina výroků, obvykle pozorování, tak abdukci můžeme chápat jako úlohu najít množinu hypotéz *H* ve vymezeném prostoru A abdukovatelných tvrzení, pro kterou platí, že

1. G je důsledkem $B \cup H$.
2. $B \cup H$ není sporná, což mj. znamená prioritu a stabilitu znalostního zázemí.
3. $H \subseteq A$ a $G \cap H = \emptyset$. Tento požadavek vyjadřuje netriviálnost abduktivního závěru.
4. G není (logickým) důsledkem znalostí B samotných.
5. Neexistuje vlastní část H´ množiny H taková, že G by bylo důsledkem $B \cup H'$, což mj. znamená, že H je minimální množinou s uvedenými vlastnostmi.

Abdukci je tedy možné takto chápat jako tentativní přijetí množiny hypotéz. Racionalita je tu měřena tím, že se hledá v prostoru kandidátských hypotéz (abducibles) maximální množina přesně vymezených vlastností, jíž je explanační minimalita. Feyerabend navíc

vyžaduje, aby ona explanace byla fundovaná v tom smyslu, že není vysvětlitelná v termínech jiných explanací.

Poznámka: Není ovšem nutné, aby formule ze všech tří množin *B, G* a *A* (a tudíž i *H*) byly nutně definovány v témž jazyce, i když společná část je patrně nezbytná. Jen těžko bychom hledali rozumnou definici abdukce mezi formulemi, které jsou navzájem irelevantní.

2.5 Trochu formální pohled na dynamiku poznávání

Po více než dvě tisíciletí jsme se z logického hlediska zabývali převážně „věčnými pravdami“ a nebrali vůbec do úvahy čas. To prostě proto, že je to velmi obtížné. Navíc tomistická absolutizace aristotelské logiky po celé milénium vytvářela bariéru pro porozumění dynamice znalostí, protože klasický princip dyadické pravdivostní funkcionality pokrývá jen nepatrnou část logické povahy významu výrazů (formulí) a porozumění logické korektnosti argumentů. Odpověď na otázku, co je to *logicky korektní argument*, je ale stále předmětem zkoumání. Filozofická reformulace musí samozřejmě vzít v úvahu na jedné straně jazykové užití oněch konceptů a na straně druhé cíle, k nimž mají být relevantní důvody hledány.

Usuzování není jen deduktivní proces, dokonce ne především, a to i tehdy, když jde o postupy, které bychom stále ještě byli ochotni označit jako racionální, tj. dobře podložené argumenty. V posledních dvou desetiletích začala být intenzivně studována logika nededuktivních úsudků (viz např. P. Gärdenfors [3], D. Gabbay et al. [4]. Ve slovenštině vyšla nedávno kniha J. Šefránka [5], v češtině je připravována publikace P. Jirků [6].)

Vlastnosti dynamiky znalostí lze formulovat v termínech expanze, kontrakce a revize teorií. Když $r(T, \alpha)$ označuje revizi teorie T formulí α; $e\{T, \alpha)$ označuje expanzi teorie T pomocí formule α, tak tzv. Gärdenforsovy postuláty racionality [3] lze formulovat následovně:

1. $r(T, \alpha) = Cn\,(r(T, \alpha))$
2. $\alpha \in r(T, \alpha)$
3. $r(T, \alpha) = e\,(T, \alpha)$
4. Jestliže $T \cup \{\alpha\}$ je bezesporná tak potom $e\,(T, \alpha) = r(T, \alpha)$.
5. $r(T, \alpha)$ je sporná právě tehdy, když $\neg\, \alpha \in Cn\,(\emptyset)$. Jinými slovy, revize pomocí kontradikce je absurdní.
6. Jestliže $Cn\,(\alpha) = Cn\,(\beta)$, potom $r(T, \alpha) = r(T, \beta)$, tj. revize téže teorie pomocí logicky ekvivalentních formulí jsou stejné.

Tyto postuláty jsou společné pro většinu dobře známých systémů nemonotónní inference. Další dva postuláty, přestože se zdají být velmi přirozené, již nejsou splněny ani v těch nejznámějších systémech jako třeba v Reiterově logice defaultů [7]. Jejich formulace navíc závisí na struktuře užitého jazyka

7. $r(T, \alpha \wedge \beta) \subseteq e\,(r(T, \alpha), \beta)$
8. $e\,(r(T, \alpha), \beta) \subseteq r(T, \alpha \wedge \beta)$ pokud $\neg\, \beta \notin r\,(T, \alpha)$

Ne všechny „přirozené“ vlastností ovšem mohou být zařazeny mezi postuláty racionality. Zde jsou příklady vlastností, které z dobrých důvodů *nemohou* být kandidáty na to být mezi obecnými axiomy pro systémy nemonotónní inference i když se na první pohled zdají být velmi intuitivními a přirozenými:

- *Jestliže $\beta \in T$, potom $\beta \notin r(T, \alpha)$ nebo $\neg \beta \in r(T, \alpha)$. Jinými slovy, revize teorie T pomocí formule α obsahuje maximální konzistentní podmnožinu množiny T, která není v kontradikci s formulí α.*
- *Jestliže $T1 \subseteq T2$, potom revize „menší“ teorie prostřednictvím formule α je rovněž částí revize „větší“ teorie, tj. $r(T1, \alpha) \subseteq r(T2, \alpha)$.*
- *Revize teorie T pomocí logicky silnější formule je částí revize téže teorie pomocí logicky slabší formule. (Pozn. Formuli α nazýváme logicky slabší než formuli β, když formule $\alpha \rightarrow \beta \in Cn(\emptyset)$.)*

Nejznámějšími příklady teorií nemonotónní inference jsou: *defaultové logiky* (Reiter 1980), *omezení* (*circumscription*) (McCarthy 1969, Lifshitz 1984, Etherington 1985, Przymusinski 1988), *autoepistemické logiky* (Nute 1980, Delgrande 1986, Pollock 1986), *kondicionální logiky* (Stalnacker 1968, Adams 1975, Loui 1987), *hierarchie dědičnosti* (Roberts a Goldstein 1977, Bobrow a Winograd 1979, Hayes 1979, Touretzky 1984, Horty a Thomason 1989, Delgrande 1987) a další systémy.

Vlastnosti racionálního usuzování, které jsme popisovali až dosud, jsou založeny na teoretických konceptech popisovaných výše. Aristoteles nás před více než dvěma tisíci roky naučil, že jedny výroky souvisejí s jinými, jsou důsledky jiných, dokonce, že celé množiny výroků (teorie) souvisejí s jinými množinami výroků. Jsou dva způsoby, jak se dovědět něco o pravdivosti výroků a o korektnosti úsudků: *evidence* a *rozum*. Máme dát přednost evidenci nebo rozumu? Logika straní rozumu, ale jen do určité míry. Zdá se, že logické důkazy jsou redukovány na posloupnosti evidencí. Pro člověka to však má své meze. Vždy jsme schopni jen nepatrného počtu evidencí. Například iterované aplikace pravidel odvozování jsou ústupkem, který sice činíme ve prospěch rozumu (dlouhé důkazy rozhodně nejsou evidentní), ale vždy se ochotně vracíme k evidencím. V této souvislosti se naskýtá otázka, zda stejné meze má i stroj. Ať tak či onak, jako hledači důvodů a důsledků narážíme jednak na svá vlastní omezení, jednak na skutečnost, že svět se neustále mění. Logika odkázaná nám Aristotelem a v minulém století rozvíjená převážně v prostředí matematiky a zahleděná právě do základů matematiky, vyzdvihnuvší statickou stránku relace odvoditelnosti, se v současné době snaží postihnout i dynamiku usuzování, tj. nalézt racionální prostředky charakterizující změny epistémických stavů, jak jsme to viděli v předchozím odstavci.

Jedním z centrálních pojmů, o které v logice jde, je pojem *kontradikce*, česky *protiřečení.* V tradičním statickém pohledu z kontradikce vyplývá cokoliv. Kontradikce je explozivní, platí princip *ex contradictione sequitur quodlibet*. Ne tak je tomu ale např. v běžném každodenním usuzování nebo při usuzování v měnícím se světě.

Z různých podnětů, jako je třeba invaze počítačů do našeho života, se na pořad dostává otázka, co z logiky lze přenechat stroji (automatické dokazování, strojové učení, induktivní logické programy). Vznikají rozmanité kalkuly. Jak se v nich organizovat? Kdy je vhodné, racionální, použít ten, kdy jiný kalkul a proč? Je proto vhodné zkoumat velmi obecné vlastnosti logických systémů, které nám umožní porozumět inferenčním

postupům, tj. tomu, co vlastně děláme, když odvozujeme. Korektnost odvození či zachovávání pravdivosti je jistě příkladem takové důležité vlastnosti. Mnohé logické systémy se těší dalším vlastnostem. Deduktivní systémy oplývají mnoha z nich [14]. Patří sem např.

- *perzistence* – tvrzení zachovávají platnost i po přidání dalších formulí
- *koherence* – libovolná formule nemůže být pravdivá i nepravdivá v témž modelu
- *determinovanost* – v úplném modelu je pravdivostní hodnota formule určena jednoznačně
- *spolehlivost* – pravdivostní hodnota formule v částečném modelu je zachována i v úplném modelu.

Při jiných než deduktivních odvozeních na některé z takových „pěkných“ vlastností musíme rezignovat. Existence některých z nich je zas jinými vlastnostmi vynucena. Tak například je-li systém perzistentní a determinovaný, je i spolehlivý. Klade se však otázka, na které ze zmíněných vlastností můžeme rezignovat, abychom naše úsudky mohli stále ještě označit za racionální. Jaká že jsou tedy kritéria racionality? Něco jsme už naznačili, ale je třeba říci, že podobnými otázkami se zabývali velcí myslitelé již mnohem dříve, např. Bernard Bolzano.

2.6 Jak lidé provádějí úsudky

Johnson-Laird v [9] tvrdí, že logika sama nestačí, mnoho logických důsledků je banálních a žádný zdravý člověk (kromě logiků) netouží takové důsledky odvozovat. Proto je zajímavým podnětem pro porozumění racionálnímu usuzování či racionálnímu argumentování i empirické studium toho, jak usuzují lidé v běžných situacích.

V každodenním životě odvozujeme *užitečné* závěry a ta užitečnost snad tkví v tom, že aktualizujeme či „visualisujeme“ informaci obsaženou v premisách a úsporně vyjadřujeme něco, co v premisách není řečeno explicitně. Johnson-Laird[2] tvrdí, že logická kompetence není věcí pouze formálních pravidel, ale spíše mentálních modelů. Lidé si vytvářejí modely situací popsaných premisami na základě jazykových a obecných znalostí. Mentální modely vyjadřují sémantickou informaci úsporným způsobem a činí explicitním to, co v premisách není vyjádřeno přímo. Je samozřejmě snazší dělat inference na základě jednoho modelu než více modelů. Chybné závěry jsou obvykle důsledky toho, že vezmeme v úvahu jen některé modely ze všech možných. Ale to, jak jsme viděli, nemusí být nahlíženo vždy jen jako závada (srovnej např. preference na modelech).

„When people reason deductively, they start with some information – either evidence of the senses or a verbal description – and they assess whether a given conclusion follows validly from this information. In real life, there is often no given conclusion, and so they generate a conclusion for themselves. Logic alone is insufficient to characterize intelligent reasoning in this case, because any set of premises yield an infinite number of valid conclusions. Most of them are banal, such as the conjunction of a premise with itself, and no sane individual, apart from a logician, would dream of drawing such conclusions. Hence, when individuals make a deduction in daily life, they must be guided by more than logic. They draw useful conclusions. The evidence suggests that they tend to maintain the information conveyed by the premises, to re-express it more

[2] Johnson-Laird je původem psycholog, takže přece jen malý návrat k psychologickým aspektům usuzování? Zde ale jde spíše než o psychologii argumentace o podněty pro formální logické zkoumání

parsimoniously, and to establish something not directly asserted in a premise. If nothing meets these constraints, they declare that there is no valid conclusion." [9]

Empirické zkoumání toho, jak lidé provádějí dedukce, může vést k novým podnětům pro rozvoj logických základů racionálního usuzování, může např. přivést k zajímavým relacím na modelech, které budou v pozadí nemonotónních operací konsekvence. Jaké obecné vlastnosti by měla racionální relace důsledku splňovat je diskusní. Gabbay [4] se domnívá, že relace důsledku, která si bude činit nárok na označení racionální, by měla být alespoň reflexivní, opatrně monotónní a tranzitivní. Už teď však lze předložit argumenty, které ukazují, že je to patrně příliš silný požadavek.

Poděkování: Tento text byl zpracován s částečnou podporou grantu GA ČR 401/01/0283

Literatura

[1] Jirků, P., Kelemen, J.: Kapitoly z kognitivní vědy. Racionalita z hlediska chování, jazyka a logiky. Praha: FIS VŠE 1996.
[2] Hájek, P.: *Gödelův důkaz existence Boha*. V Jaroslav Malina – Jan Novotný (ed.) Kurt Gödel. Brno: Nakladatelství a vydavatelství Malina 1996.
[3] Gärdenfors, P.: *Knowledge in Flux. Modeling the Dynamics of Epistemic States*. MIT Press 1988.
[4] Gabbay, D., Hogger, C. J. Robinson, J. A. (eds.): *Handbook of Logic in Artificial Intelligence and Logic Programming*., Vol 1-4. Oxford: Clarendon Press 1994.
[5] Šefránek, J.: *Inteligencia jako výpočet*. Bratislava: Nakladatelství IRIS 2000.
[6] Jirků, P.: Teorie racionáního usuzování. (to appear).
[7] Reiter, R.: A Logic for Default Reasoning. *Artificial Intelligence* **13** (1980) 81-132.
[8] Josephson, J. R.: *On the "Logical Form" of Abduction*. Web file. http://www.cis.ohio-state.edu/lair/TechReports/logical-form-working.ps
[9] Johnon-Laird, P. N., Byrne, R. M. J.: *Deduction*. *Essays in Cognitive Psychology*. Howe and London: Lawrence Erlbaum Associates Ltd. 1991.
[10] Johnson-Laird, P. N.: *Mental Models*. Cambridge, MA: Harvard University Press 1983.
[11] Chaitin, G. J.: *The unknowable*. Singapore: Springer 1999.
[12] Jirků, P.: *Possibile et impossibile*. Horizonty vědění. Liberec: Scholé filosofia 2000, 19-30.
[13] Jirků, P.: *How to understand negatives*. Yearbook of Logic, Praha: Filosofia 2001.
[14] Kneale W., Kneale, M.: *The Development of Logic*. Oxford: Clarendon Press 1962.
[15] Smullyan, R. M.: *Gödel's Incompleteness Theorems*. New York – Oxford: Oxford University Press 1992.

Pokus o klasifikáciu optických ilúzií (od paobrazov k nemožným obrazcom)

Ján Rybár[1]

Abstrakt. Skúmanie optických ilúzií je nesporne veľmi atraktívna a zároveň efektívna cesta ako lepšie pochopiť procesy vizuálneho vnímania. Veľmi dôležitým systémovým predpokladom takéhoto skúmania je klasifikácia týchto ilúzií (resp. klasifikácia techník, na základe ktorých sa vytvárajú tieto ilúzie). A práve tomuto problému je venovaný náš článok.

1 Úvod

V súčasnosti je k dispozícií obrovské množstvo techník na demonštrovanie najrôznejších optických ilúzií (môžeme hovoriť o stovkách a ich počet neustále rastie). O tom svedčí aj množstvo webovských stránok venovaných optickým ilúziám. Ak chceme vniesť do toho nejaký poriadok (alebo ak chcete, vedeckejší prístup) prvé, čo treba urobiť, je pokúsiť sa o klasifikáciu týchto ilúzií. Niektoré takéto pokusy boli už urobené. Medzi najzaujímavejšie patrí klasifikácia R. Gregoryho. V podstate delí ilúzie podľa dvoch základných kritérií. Prvé kritérium spočíva v tom o aký druh chýb pri videní ide. Podľa toho môžeme deliť ilúzie na: dvojznačné obrazce (napríklad Neckerova kocka, káčer-zajac a mnohé ďalšie), skreslenia (všetky ilúzie, ktoré sa týkajú porušenia konštantnej veľkosti, tvaru a jasu), paradoxy (napríklad Penrosove schody) a fikcie (napríklad Kaniszov trojuholník). Druhým kritériom klasifikácie ilúzií je delenie podľa toho, čo ich spôsobuje. Podľa tohto kritéria delíme ilúzie na: fyzikálne (napríklad palica ponorená do vody sa zdá nalomená), fyziologické (rôzne paobrazy a afterefekty), kognitívne (napríklad dvojznačné obrazce), a ktoré súvisia s geštaltistickými princípmi (princíp blízkosti, podobnosti, symetrie, atď.) ([5], str. 248-249).[2]

V hre sú aj ďalšie kritéria, ak sa vstupné senzorické dáta menia, ale obrazce nejakým spôsobom pretrvávajú, hovoríme o fyziologických ilúziách (napríklad paobrazy).[3] A naopak, ak vstupné dáta sa nemenia, ale obrazce napriek tomu áno, hovoríme o kognitívnych ilúziách (dvojznačné obrazce, fikcie, skreslenia – veľkosti, tvaru, jasu a tiež tzv. nemožné obrazce). Gregory poznamenáva, že fyziologické ilúzie sú o mozgu, kognitívne o mysli.

V klasifikácii, ktorú používame, sa pridržiavame posledných dvoch kritérií, teda či sa menia vstupné senzorické dáta alebo nie a aký to má vplyv na obrazce v procese

---

[1] FMFI UK – Katedra aplikovanej informatiky, Centrum kognitívnych vied, Mlynská dolina, 842 48 Bratislava, E-mail: rybar@fmph.uniba.sk

[2] Zaujímavá je aj kategorizácia optických ilúzií podľa obsiahnutých postupov, ktorej autorom je J. Ninio [8]. Členením predpokladov optických ilúzií aplikovaných v oblasti architektúry sa zaoberá R. Šikl a M. Šimeček [12].

[3] Aby sme sa vyhli prípadným nedorozumeniam, upozorňujeme, že termín fyziologické ilúzie používame v tomto článku len v tomto presne zadefinovanom zmysle. Samozrejme, že v širšom zmysle slova sú všetky ilúzie fyziologické, pretože sú vždy spojené s činnosťou oka a mozgu.

vnímania. Tieto základné kritériá kombinujeme s klasifikovaním podľa druhu chýb, ktoré sa vyskytujú v optických ilúziách.

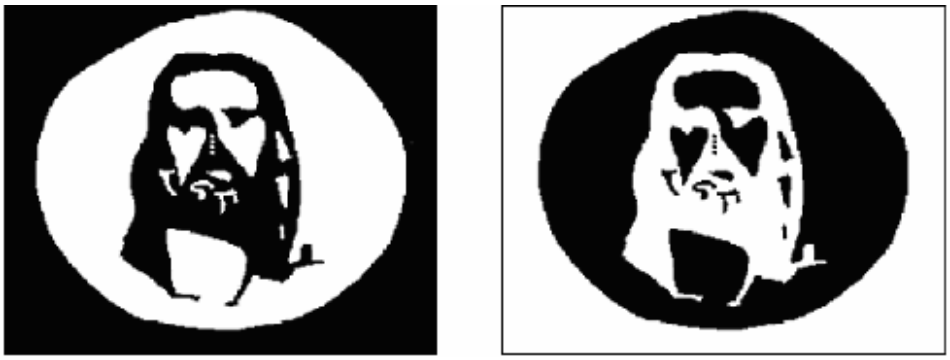
2 Fyziologické ilúzie

2.1 Paobrazy

Typickým predstaviteľom fyziologických ilúzií sú tzv. paobrazy (after-images). O existencii týchto fenoménov sa môžete veľmi ľahko presvedčiť, ak sa pozriete na niekoľko sekúnd do ostrého svetla, napríklad žiarovky, a potom zatvoríte oči, alebo premiestnite pohľad na bielu plochu, uvidíte plávajúcu žiaru.

Určite však najslávnejšou a najznámejšou ilúziou v tejto kategórii je ilúzia bradatého muža (na webovských stránkach ju môžete nájsť aj pod názvom *four dots*). Je to veľmi efektný príklad ilúzie. Ak sa budete pozerať 30 sekúnd na isté škvrny, uprostred ktorých sú štyri bodky (pozri pravú stranu obrázka č. 1), a potom zatvoríte oči, na mentálnej obrazovke, alebo ak premiestnite pohľad na bielu plochu, uvidíte v svetelnom kruhu hlavu bradatého muža.

Na prvý pohľad sa môže zdať, že táto ilúzia je zaujímavou ilustráciou jednej z mnohých záhad obsiahnutých v defaulte vizuálnej mozgovej kôry (resp. ďalších častí mozgu). Cestu k vysvetleniu tejto záhady treba však primárne hľadať už v tom, čo sa v tejto súvislosti odohráva na sietnici. Pri vnímaní čiernobieleho obrazca bunky na citlivé svetlo sa unavia, a to vlastne spôsobuje, že vidíme negatívny after-image. Tieto experimenty sú veľmi dobrými príkladmi na demonštráciu toho ako na sietnici pracujú bunky citlivé na svetlo.

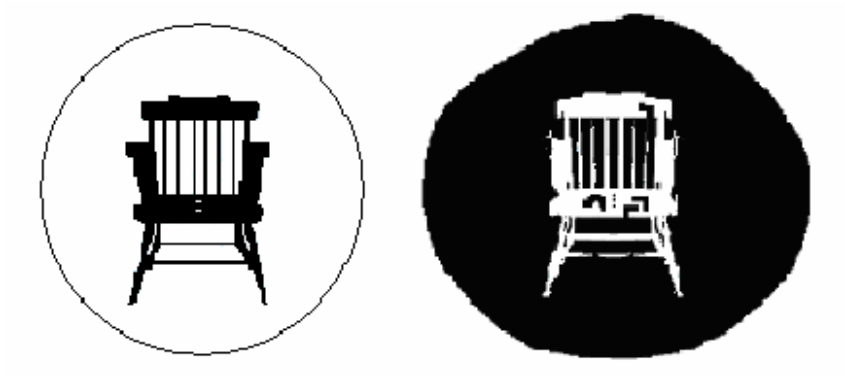


Obrázok 1. Originál a negatív.

Na základe toho sme si vytvorili hypotézu, že konštruktéri tejto ilúzie nám vlastne svojím spôsobom predkladajú negatív hlavy bradatého muža. V zásade je to tak, ale v hre sú ešte ďalšie záležitosti, pretože, ak urobíme počítačový negatív negatívu (pozri ľavú stranu obrázka č. 1) nedostaneme dokonalú hlavu bradatého muža. Tvorcovia ilúzie, aby sme celkom ľahko neprišli na to, že je to jednoducho negatív, urobili ešte niekoľko zásahov do tohto negatívu – kruh, v ktorom je negatív bradatého muža, je trochu nepravidelný, črty tváre nie sú celkom ostré, sú tam pridané určité biele škvrny. Nič z toho v after-image nevidíme, to znamená, že pri konštruovaní tejto ilúzie, okrem čiernobieleho

efektu, boli využité ešte aj ďalšie záležitosti súvisiace s geštaltistickými zákonmi vnímania (predovšetkým zákon pregnantnosti).

Na jednom seminári, keď sme demonštrovali tieto skutočnosti, bola položená otázka, či sa objaví ten istý fenomén, ak na obrázku bude napríklad stolička. Samozrejme, že to funguje aj so stoličkou, alebo s akýmkoľvek čiernobielym objektom. Dokonca v súlade s Heringovými objavmi to funguje, ak použijeme nielen čiernobiely, ale aj žltomodrý alebo červenozelený negatív. Jednoducho, ak použijeme tzv. komplementárne farby. Na základe toho bolo vytvorených mnoho after-imagových ilúzií, v ktorých sa pracuje s farbami štátnych zástav (zastáva USA, Anglicka, Nemecka, Japonska a podobne).



Obrázok 2. Originál stoličky a zdeformovaná verzia.

Aj v prípade stoličky platia tie isté záležitosti ako v prípade after-imagu bradatého muža. Keď sme do negatívneho obrázka stoličky pridali čierne škvrny, trochu sme rozrušili jej ostré línie (pozri pravú stranu obrázku č. 2), napriek tomu after-imagová stolička bola bez škvŕn, jej línie boli ostré a bola v dokonalom svetelnom kruhu. Mimochodom, štyri bodky uprostred, podľa ktorých sa zvykne ilúzia bradatého muža volať (na našom obrázku stoličky sme použili len tri bodky), nemajú v ilúziách tohto druhu veľký význam, je to len ďalšia odpútavacia finta.

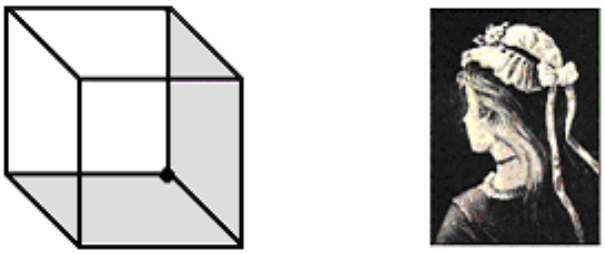
3 Kognitívne ilúzie

Ako sme už uviedli, základným kritériom pre určenie kognitívnych ilúzií je fakt, že hoci sú vstupné dáta stále, napriek tomu sa vnímané obrazce menia. Tieto ilúzie sú oproti fyziologickým ilúziám viac spojené s vysokoúrovňovými kognitívnymi procesmi. Kognitívne ilúzie podľa druhu chýb môžeme deliť na niekoľko podkategórií: dvojznačné obrazce, fiktívne obrazce, skresľujúce (veľkosť, tvar a jas) a nemožné obrazce.

3.1 Dvojznačné obrazce

Ide o figúry, ktorých vnímanie je spojené s dvomi, niekedy aj viacerými možnosťami. Nezainteresovanému sa zdá, že prepínanie tých možnosti je náhodné, nie je to celkom tak. V prípade Neckerovej kocky (ľavá strana obr. č. 3) je čierny bod niekedy vo vnútri a inokedy z vonkajšej strany. Náš vizuálny systém v tomto prípade dokáže cieľavedome

preklikávať. Nikdy však nemôžeme vidieť obidve orientácie naraz. Podobne je to aj v prípade populárnej dvojznačnej figúry (pravá strana obr. č. 3), ktorá sa zvykne nazývať *svokra-nevesta.* Vstupný obrázok sa nemení, ale náš vizuálny systém ho interpretuje dvomi spôsobmi. Závisí to od toho, kde sa pozeráme. Ak sa pozeráme na oko, uvidíme svokru, ak viac na ľavú stranu obrázku, uvidíme nevestu.



Obrázok 3. Neckerova kocka a svokra-nevesta.

Dvojznačnými (resp. mnohoznačnými) figúrami sa inšpirovalo mnoho slávnych výtvarných umelcov. Nadchol sa nimi aj známy surrealistický maliar Salvador Dalí, ktorý tento druh ilúzií zakomponoval do mnohých svojich obrazov. Najznámejšie z nich sú *Paranoická figúra* (na obraze môžeme vidieť tvár ležiacu na piesku alebo skupinu sediacich ľudí), *Obraz mizne* (môžeme tu vidieť hlavu fúzatého a bradatého muža z profilu, alebo čítajúcu ženu) a ďalšie.

Od konca 19. storočia, keď sa psychológovia začali intenzívne zaujímať o optické ilúzie, bolo skonštruovaných veľmi veľa dvojznačných figúr. Nedávno sa ich J. R. Blockovi do knihy *Seeing double* podarilo nazbierať viac než dvesto.

3.2 Fiktívne ilúzie

Tento druh ilúzií tiež svedčí proti koncepciám priamej percepcie (vnímané objekty nie sú v senzorických vstupoch). Na Ehrlichovom obrázku (vľavo na obr. č. 4) nie sú nakreslené žiadne biele kruhy, a predsa ich tam jasne vidíme.

Najslávnejšou a najefektnejšou ilúziou v tejto oblasti je Kaniszov trojuholník (stredný a pravý objekt na obr. č. 4). Biely trojuholník, ktorý vidíme, je fikciou. Ide o top-downový proces, náš vizuálny systém si obrázok interpretuje tak, že konštruuje trojuholník, ktorý prekrýva tri tmavé objekty. Trojuholník má výraznejšiu bielu farbu než pozadie. Je zjavné, že sú tu v hre aj geštaltistické princípy – princíp uzatvorenosti, dobrých tvarov, atď. Ilúzia fiktívneho trojuholníka sa objaví aj vtedy, ak pod ním nie je nakreslený ďalší akoby prekrytý trojuholník. Úloha tohto trojuholníka spočíva iba v zosilnení účinku ilúzie.

Obrázok 4. Ehrlichova ilúzia a Kaniszov trojuholník.

3.3 Skresľujúce ilúzie

Porušenie konštantnosti veľkosti. Prvá skupina ilúzií, ktorou sa tu budeme zaoberať, sa týka fenoménu vnímania konštantnej veľkosti objektov. Ide o to, že napriek tomu, že veľkosť objektu na našej sietnici sa mení podľa toho ako je predmet vzdialený od nás (ak je objekt bližšie k nám je väčší, ak je ďalej, je menší na našej sietnici), stále máme rovnakú predstavu o jeho veľkosti. Táto stálosť sa však pomocou rôznych manipulácií dá významne narušiť. Hádam najznámejším klasickým prípadom je Müller-Lyerova figúra.

Obrázok 5. Müller-Lyerova ilúzia.

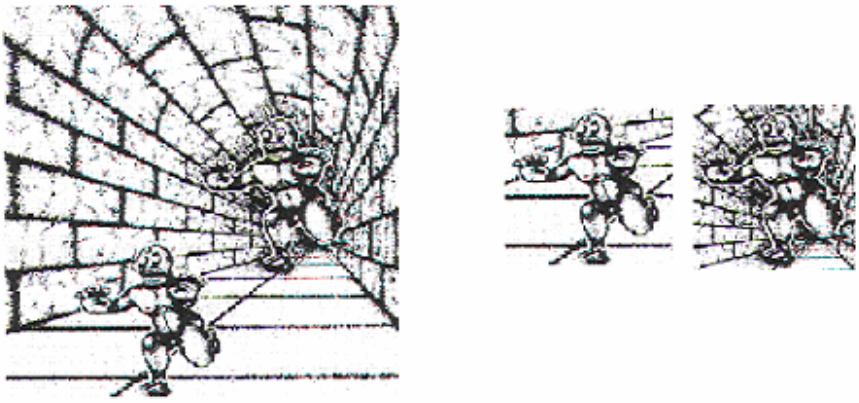
Podstata ilúzie je v tom, že ak má jedna z dvoch rovnakých úsečiek kótovanie (krídelká) dovnútra a druhá von, úsečka s vonkajším kótovaním (s vonkajšími krídelkami) sa bude zdať väčšia. Napríklad pri pokusoch, v ktorých sa porovnáva štandardná úsečka (bez krídelok) s úsečkou s krídelkami von, sa zistilo, že v odhade veľkosti sa mýlime v priemere o 10% [3].

Zaujímavé je, že ilúzia ostáva, aj keď vieme, že je to ilúzia. Tento poznatok, táto vedomosť, ju neodstraňuje. Teda v tomto zmysle percepcia nie je informačne otvorená voči predchádzajúcim skúsenostiam. Zdá sa, že nie je penetrovaná kogníciou a je informačne uzatvorená, má modulárnu povahu ([10], str. 35-40).

Často si hneď neuvedomíme, že v súvislosti s Müller-Lyerovou figúrou ide vlastne o dve ilúzie. Druhá ilúzia spočíva vo fenoméne, že dvojrozmerné útvary vidíme ako trojrozmerné. Ľavú úsečku vidíme ako vnútorný roh nejakej miestnosti a pravú ako vonkajší roh nejakej stavby. Problému vnímania dvojrozmerných útvarov ako trojrozmerných sa venujeme viac v záverečnej časti nášho článku. V tejto súvislosti sa vedú

rôzne diskusie nakoľko je táto ilúzia produktom kultúrneho prostredia a nakoľko je vrodená.[4]

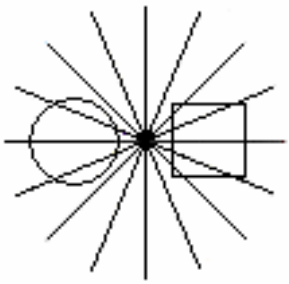
Konštantnosť veľkosti môžeme narušiť aj pozadím, v ktorom sú línie v perspektíve. Ukážkovým príkladom je Shepardova ilúzia (obr. č. 6). Pri porovnávaní utekajúcich figúrok vidíme, že sú identické. Napriek tomu figúrka umiestnená na vhodnom pozadí (v tomto prípade v perspektíve zbiehajúcich sa línií nakresleného tunelu) sa zdá oveľa väčšia.



Obrázok 6. Shepardova ilúzia.

Z hľadiska o aký druh chyby ide, obidve ilúzie (Müller-Lyerova a Shepardova) evidentne patria do kategórie skresleného, deformovaného vnímania.

Porušenie konštantnosti tvaru. Podobne je to aj s konštantnosťou tvaru. V závislosti od toho, či sa pozeráme na predmet spredu alebo zboku je na sietnicu premietnutý iný tvar, ale to nijako neohrozuje našu predstavu o stálom tvare predmetu. Avšak pomocou pozadia (napríklad líniami usporiadanými do hviezdice) vieme narušiť túto konštantnosť. Bolo vymyslených mnoho ilúzií tohto druhu, niektoré už v 19. storočí. Ich autormi boli fyzici, ktorí sa zaoberali skresleniami optických prístrojov. Uveďme aspoň jednu z nich (pozri obr. č. 7).

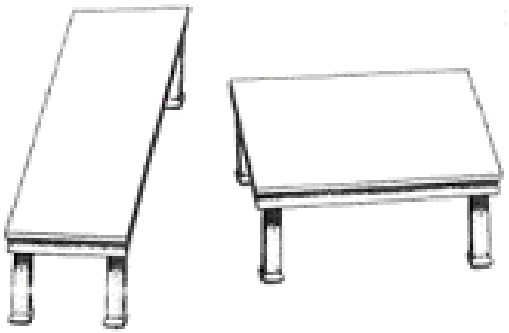


Obrázok 7. Orbisonova ilúzia.

[4] V niektorých medzi-kulturálnych porovnávacích výskumoch sa ukázalo, že existujú významné rozdiely v citlivosti na Müller-Lyerovu figúru ([8], str. 150-151).

V odbornej literatúre je popísaný zaujímavý prípad S. B. Išlo o muža, ktorý bol slepý od narodenia až do stredného veku. Po transplantácii rohovky začal vidieť. Jeho videnie bolo spojené s mnohými ťažkosťami a zvláštnosťami a trpel depresívnymi náladami. Mal napríklad veľký problém s tým, že predmety z rôznych strán vyzerajú rôzne. Vizuálny svet sa mu nezdal taký dokonalý ako očakával. Zle odhadoval vzdialeností vecí, ktoré nepoznal z dotyku, podobné ťažkosti mal aj s kreslením vecí, ktoré nepoznal z dotyku. Mal problémy s čítaním malých písmen, pretože tie sa v slepeckej škole neučili [5]. Okrem iného S. B. nebol veľmi senzitívny na ilúzie Orbisonovho typu.

Konštruktéri podivných fenoménov radi využívajú aj ilúzie spojené s vnímaním vertikály a horizontály. Je známe, že vzdialenosti orientované zdola hore (alebo zhora dole) sa nám zdajú väčšie než vzdialenosti v orientácii zľava doprava (alebo sprava doľava).[5] Čiastočne aj o tento poznatok sa opiera konštrukcia Shepardovho stola. Ide o ilúziu týkajúcu sa tvaru a veľkosti dosiek stolov, ktoré hoci nevyzerajú rovnaké, napriek tomu sú toho istého tvaru a veľkosti (obr. č. 8). Avšak preceňovanie vertikály ani zďaleka nestačí na vysvetlenie ilúzie, pretože, keď obrázok otočíme o 90 stupňov, ilúzia sa síce trochu zoslabí, ale nezmizne (stoly stále vyzerajú nerovnaké). To znamená, že pri konštruovaní tejto ilúzie sú v hre aj ďalšie záležitosti. Veľmi dôležitým faktorom je, že hrany a nohy stolov sú v perspektíve, čo vytvára hĺbku a predlžuje jeden stôl do dĺžky a druhý do šírky. V hre sú pravdepodobne aj ďalšie faktory monokulárneho videnia hĺbky. V každom prípade ide tu o využitie viacerých ilúziotvorných prvkov v jednom obrazci [8].

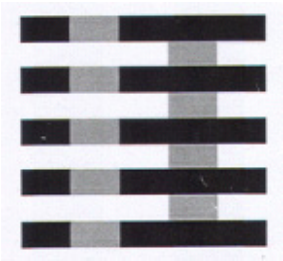


Obrázok 8. Shepardov stôl.

Porušenie konštantnosti jasu. Existuje veľké množstvo efektných ilúzií tohto druhu. Na ilustráciu sme vybrali Whitovu ilúziu, ktorá patrí medzi najjednoduchšie. Na obrázku č. 9 je v obidvoch stĺpcoch rovnaká šedá. Jasnosť šedého políčka sa mení v súvislosti s pozadím. Na tmavom pozadí sa tá istá šedá farba javí svetlejšia a naopak na svetlom pozadí sa javí tmavšia.

Aj táto ilúzia patrí do kategórie skresleného vnímania. Veľmi zaujímavé analýzy o fenoménoch s porušovaním jasnosti môžeme nájsť v štúdii J. Kremláčka [7].

[5] Evolučno-psychologické vysvetlenie tejto ilúzie hovorí, že predimenzované videnie vertikály bolo výhodné pre lovca-zberača.



Obrázok 9. Whitova ilúzia.

3.4 Nemožné obrazce

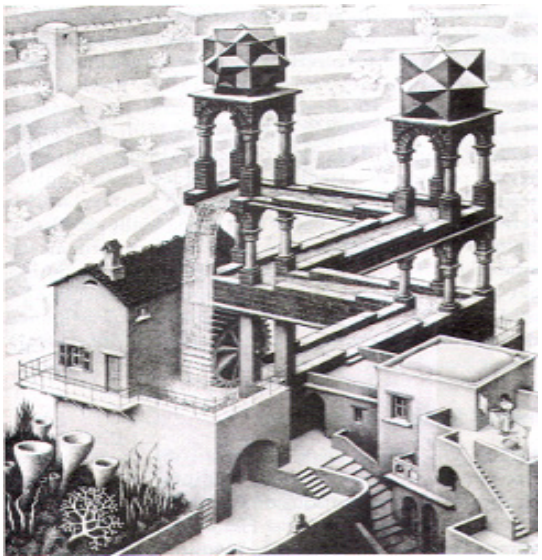
Medzi najznámejšie ilúzie tohto druhu patrí trojzubec (trident), ktorý má veľa modifikácií (napríklad Riman, slon a ďalšie). Veľmi efektnou ilúziou sú Penrosove schody (ich animovanú podobu môžeme vidieť napríklad v Zimbardovom televíznom kurze psychológie). Tvorcovia ilúzií tohto druhu boli často slávni maliari (Dalí, Escher, Magritte a iní).

Nemožné objekty majú zvláštnu povahu, môžeme ich nakresliť v dvojrozmernej rovine, ale nemôžeme ich fyzicky skonštruovať v trojrozmernom priestore. Holandský grafik M. C. Escher bol skutočne majstrom vytvárania podivných trojrozmerných efektov (ilúzií) v dvojrozmernej rovine, bol majstrom v mýlení našich schopností vnímať trojrozmerné objekty. Jeho grafiky pôsobia ako akýsi vizuálny labyrint, ktorého zmysel nechápeme, a z ktorého sa nedá dostať von. Nie je to však až také beznádejné. Odpoveď ako sa zorientovať v labyrinte Escherových trikov nám môže poskytnúť práve psychológia vnímania.

Znamenitým príkladom nemožného obrazca je Escherova grafika s názvom *Vodopád* (obr. č. 10). Na grafike je prezentovaný mlyn s vodným náhonom a dvomi vežami. Ak sa budeme na obrázok pozerať analyticky (kúsok po kúsku 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 holistickým prístupom.

Na prvý pohľad často vyzerajú Escherove nemožné objekty ako veľká fantázia, keď sa však na to pozrieme cez psychológiu vnímania zistíme, že sú zostrojené podľa prísnych pravidiel, že rešpektujú pravidlá na zobrazovanie hĺbky. Na obrázku sú šikovne využité vodidlá pre monokulárne videnie hĺbky (podľa ktorých určujeme hĺbku, pretože na sietnici je dvojrozmerný obraz, trojrozmerný obraz je dielom vizuálnej mozgovej kôry).

Prvé také vodidlo je prekrývanie objektov. Hoci i len čiastočne prekryté objekty sa na dvojrozmernom obraze vždy zdajú vzdialenejšie. Druhé vodidlo súvisí s umiestnením objektov vo výške. Objekty umiestnené vyššie (nad horizontom) sa zdajú vzdialenejšie, objekty umiestnene nižšie (pod horizontom) sa zdajú bližšie. V danom prípade je využitý výškový efekt, napriek tomu, že kanál smeruje v rovine do výšky, máme dojem, že voda tečie do hĺbky. Tretie vodidlo, ktoré pomáha vytvárať hĺbku, sú konvergujúce línie, teda lineárna perspektíva.



Obrázok 10. Vodopád.

Štvrtým vodidlom 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 múriku, ktorý vodu obklopuje. Dôležitým vodidlom je povrchová štruktúra nakresleného múriku kanála aj samotnej vody, je hrubšia na začiatku kanála a jemnejšia na konci, čo tiež posilňuje ilúziu hĺbky. Ďalším vodidlom je tienenie. Tmavšie časti objektu (napríklad múriky vodného kanála) sa zdajú byť vzdialenejšie než svetlejšie oblastí. Posledným vodidlom je vzdušná perspektíva. Obrazy objektov, ktoré sú nejasnejšie, sa zdajú vzdialenejšie, než objekty s výraznými zreteľnými obrysmi. Escher práve pomocou majstrovského využívania vodidiel monokulárneho videnia hĺbky dokáže vytvoriť ilúziu nemožného, že voda tečie hore kopcom. Na začiatku sa nám mohlo zdať, že ide o záhadný výtvor autorovej výstrednej mysle. Teraz vidíme, že to nie je tak. Escherov prístup je vysoko racionálny.

4 Záver

Už i z uvedených príkladov konštrukcií optických ilúzií je zrejmé, že pokusy o ich klasifikáciu (teda pokusy o prekonanie len čisto katalógového prístupu) môžu byť veľmi užitočné a podnetné. Nútia k hľadaniu a testovaniu efektívnych nástrojov (kritérií) pre skúmanie tak rozsiahleho materiálu. Sú hľadaním základných charakteristík, ktoré sú spojené s optickými ilúziami. Tieto charakteristiky sú vlastne charakteristikami celého percepčného systému. Môžu tiež viesť k lepšiemu objasneniu toho, čo je výsledkom tzv. nízko-úrovňových a čo vysoko-úrovňových percepčných procesov. A samozrejme prispievajú k transparentnejšiemu výkladu optických ilúzií.

Poďakovanie: 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] Block, J. R.: *Seeing double*. London: Routledge 2002.
[2] Block, J.R., Yuker H. E.: *Can you believe your eyes?* London: Brunner-Routledge 1989.
[3] Francis, G., Neath, I., MacKewn, A., Goldthwaite, D.: *CogLab on CD*. London: Thomson Learning 2004.
[4] Gigerenzer, G.: I think, therefore I err. *Social Research*, 72 (1), (2005) 195-218.
[5] Gregory, R. L.: *Eye and Brain. The Psychology of Seeing.* Fifth edition. Oxford University Press 2004.
[6] Hofstadter, D. R.: *Gödel, Escher, Bach*. New York: Vintage Books 1979.
[7] Kremláček, J.: *Co prozradily optické iluze o našem zrakovém vnímání.* In: Sborník konferenčních příspěvku. Psychologický ústav Akademie věd ČR 2006.
[8] Ninio, J.: *The science of illusions*. Cornell University Press 2001.
[9] Penrose, L. S., Penrose, R.: Impossible Objects: A special type of visual illusion. *Brit. J. Psychology* **49** (1958) 31-33.
[10] Rybár, J. a kol.: *Filozofia a kognitívne vedy*. Bratislava: Iris 2002.
[11] Schattschneider, D., Emmer, M. (Eds): *M. C. Escher´s Legacy. A Centennial Celebration.* Berlin: Springer 2005.
[12] Šikl, R., Šimeček, M.: *Vnímání plné předpokladu.* Sborník konferenčních příspěvků. Psychologický ústav Akademie věd ČR 2006.

Kognitívna teória usudzovania

Ján Šefránek[1]

Abstrakt. Príspevok je pokusom o načrtnutie kognitívnej teórie usudzovania založenej na znovupremyslení McCarthyho vízie o explicitnom (a formálnom) postihnutí fenoménu skákania k záverom (jumping to conclusions).

1 Úvod

Téma usudzovania je lákavou výzvou kognitívnym vedám. Kognitívna veda totiž nevenuje primeranú pozornosť usudzovaniu, základu vyšších kognitívnych schopností. Rigorózne štandardy logiky viedli k hlbokým výsledkom založeným na pevných sémantických základoch. Tieto výsledky siahajú až po hranice formálnych a výpočtových modelov (známe výsledky o nevypočítateľnosti a nezvládnuteľnosti výpočtov). Stále však ostáva pomerne veľká a neopísaná priepasť medzi formálnymi modelmi, ktoré sú predmetom výskumov logikov a skutočným ľudským usudzovaním.

Cieľom mojej teórie je vkročiť do tejto priepasti. Vychádzam z pozorovania, že ľudia, keď usudzujú, obvykle nerobia dlhé odvodenia. Často priamo (a rýchlo) „vidia" závery. Dlhšie zdôvodnenia alebo argumentácie používajú až vtedy, keď narazia na nejaký problém (pri rýchlom usudzovaní).

Predkladaná teória má dve polohy. Dokáže opísať a charakterizovať rýchle, i keď neúplné a omylné usudzovanie. Túto polohu možno považovať za základný modus usudzovania, budem ho označovať ako skákanie k záverom. Na druhej strane, predkladaná teória umožňuje aj prechod k detailnej argumentácii, k rigoróznemu usudzovaniu. Obe črty však spája do jednotného formálneho rámca. Tento text podáva iba stručnú informáciu o projekte.

2 Motivácia a prehľad

Vplyvná (až prevratná) myšlienka skákania k záverom (jumping to conclusions) [1] významne ovplyvnila v posledných dvoch desaťročiach výskum usudzovania. Výskumníci v umelej inteligencii založenej na logike už na prelome šesťdesiatych a sedemdesiatych rokov rozpoznali, že dedukcia nepostačuje na pochopenie reprezentácie poznatkov a inteligentného[2] usudzovania [6,7,8,4].

V deväťdesiatych rokoch sa objavili mnohé hlboké výsledky v oblasti logických základov hypotetického usudzovania, ktoré významne posunuli pochopenie tejto oblasti [5], v slovenčine možno získať informácie o oblasti v [9]. Ukázalo sa, že formalizácie hypotetického usudzovania (v rozpore s očakávaniami) viedli k výpočtom ešte ťažším než sú výpočty späté s klasickou (monotónnou) dedukciou. Tieto formalizácie spĺňali striktné kritériá logickej presnosti a dôkladnosti. Platí sa však tým, že problémy, vnútorne späté so spomínanými formalizáciami, sú ťažké (z výpočtového hľadiska).

---

[1] Ústav informatiky FMFI UK, 842 15 Bratislava, E-mail: sefranek@fmph.uniba.sk

[2] Pod inteligentným usudzovaním tu máme namysli to, čomu sa anglicky hovorí common-sense reasoning. Ďalej budeme v tomto význame používať termín hypotetické usudzovanie.

Cieľom tohto článku je znovupreskúmanie myšlienky skákania k záverom. Základné formalizácie nemonotónneho usudzovania ju používali v nasledujúcom zmysle: „A reasoning agent can be said to be ‚jumping to conclusions' whenever it draws conclusions that are not deductively justified by the premises" [2].

Zmeňme[3] uhol pohľadu. Ľudia, keď rozmýšľajú, obvykle nepoužívajú dlhé odvodenia. Často priamo (a rýchlo) „vidia" závery, ktoré prijímajú (takto budeme rozumieť skákaniu k záverom). Až vtedy, keď pri skákaní k záverom narazia na nejaké problémy, prechádzajú k zdĺhavejšiemu zdôvodňovaniu, argumentácii, odvodzovaniu.

Prijmeme, že pri skákaní k záverom ľudia používajú určité *schémy*. Budeme ich ostro odlišovať od báz poznatkov (explicitne vyjadrených vetami nejakého jazyka). Skákanie k záverom, samozrejme, môže byť omylné. Preto predpokladáme, že schémy môžu byť „chybné". Iba niekedy však chyby pri skákaní k záverom vedú k problémom, ktoré si usudzujúci človek uvedomuje (k situáciám signalizujúcim, že treba usudzovať dôkladnejšie). Vtedy „prepne" na zdĺhavejšie a precíznejšie argumentovanie, odvodzovanie.

V článku budeme prezentovať formálnu teóriu kombinovaného použitia schém a dôkladného odvodzovania pri hypotetickom usudzovaní. Cieľom je zachovať myšlienku rýchleho skákania k záverom spolu s možnosťou kombinovať ho s usudzovaním, ktoré je detailné, rigorózne, úplné a založené na logickej disciplíne. Túto kombináciu stelesníme dvoma typmi agentov (usudzujúci agenti a ideálni agenti). *Usudzujúci agent* používa schémy, skáče k záverom. Keď sa dostane do ťažkostí, „konzultuje" *ideálneho agenta*, schopného detailnej logickej argumentácie. Oba typy agentov budú špecifikované v jednotnom formálnom rámci. Prvý podrobnejší opis tohto rámca možno nájsť v [11].

V nasledujúcich dvoch častiach sa budeme venovať metodologickým základom predloženého prístupu: charakterizujeme jeho observačnú bázu a teoretické konštrukty s ňou späté, sémantickú bázu. Potom opíšeme základné intuície späté s našou teóriou. Ďalej načrtneme technické detaily, na ktorých je založený súčasný stav nášho prístupu. Napokon na záver budeme diskutovať dosiahnuté výsledky a otvorené problémy.

3 Observačná báza

Čitateľ si iste všimol, že prijímame isté – pomerne silné – predpoklady (typu „Ľudia, keď rozmýšľajú, obvykle nepoužívajú dlhé odvodenia"). Tento článok nepristupuje k štúdiu usudzovania tak, ako to robia logici[4]– ich metóda (zhruba) spočíva v definovaní nejakého umelého jazyka, v definovaní jeho sémantiky, v sformulovaní nejakej teórie v tomto jazyku a v skúmaní jej vlastností. Prívlastok „kognitívna" v názve článku naznačuje, že našim cieľom je, aby tu načrtnutá teória bola empiricky falzifikovateľná (predpoklady, hypotézy, o ktoré sa opierame, sú empirickej povahy). Preto naším prvým cieľom je vymedziť, aký typ pozorovaní považujeme za relevantný pri preverovaní našej teórie. Inými slovami, vymedzíme *observačnú bázu.*[5]

Observačnou bázou našej teórie sú (v prvom pláne) *jazykové výrazy*, ktoré ľudia používajú pri *argumentácii, zdôvodňovaní* svojich názorov, pri polemike. Môže ísť o výrazy prirodzených jazykov, ale aj – viac či menej umelých – vedeckých jazykov. Pri

[3] Prechádzam k plurálu – verím, že prirodzeným spôsobom.
[4] Ani tak ako psychológovia, treba dodať.
[5] V tomto článku nám však nejde o preverovanie navrhnutej teórie.

pozorovaní argumentácie však narážame na subtílnejšie vrstvy a vzťahy. Argumentácia (formulovaná v jazyku) býva spätá s nejakým správaním, riešením problémov. A naopak, k niektorým typom *správania*, k *riešeniu problémov*, *odpovedaniu na otázky* sa dá priraďovať argumentácia, na ktorej môžu byť založené. Našu observačnú bázu musíme rozšíriť o doménu správaní, ktoré možno podložiť nejakou argumentáciou. Iste si tým skomplikujeme život, ale nemôžeme sa tomu vyhnúť.

V literatúre o kognitívnych vedách sa zvyknú analyzovať príklady mentálnych procesov ako odvodenie myšlienky „John swims“ z myšlienky „John swims and Mary drinks“ [3]. Takéto odvodenia však ľudia nerobia (pokiaľ nie sú na cvičení z výrokovej logiky) – to je ďalšia z hypotéz, ktoré prijímam. Dve dôležité poznámky k príkladu z Fodora:

1. mentálne procesy netvoria našu observačnú bázu, mentálne procesy nás zaujímajú iba vtedy, keď sa prejavujú (alebo môžu prejavovať) pomocou nejakých jazykových výrazov – vtedy však na ne môžeme zabudnúť;
2. odvodenia (typu „z A a B vyplýva A“), ktoré ľudia obvykle, ak sú normálni, nerobia (prinajmenšom na úrovni jazyka), ale stali sa zaujímavými z hľadiska logiky, naznačujú, že by sme mali rozmýšľať o ďalšom rozšírení našej observačnej bázy.

Samotná úroveň jazykových výrazov používaných pri argumentácii nepostačuje totiž na pochopenie argumentácie, zdôvodňovania, odvodzovania. Tieto procesy môžeme pochopiť až vtedy, keď dokážeme rozlišovať, kedy sú korektné a kedy nekorektné (až týmto rozlíšením sa dostaneme k podstate).

4 Sémantická báza

Kogníciu možno charakterizovať ako interaktívny, evolutívny a nedeterministický výpočet [12, 10]. My si tu všimneme interaktívnosť. Človek rozmýšľa tak, že generuje nejaké myšlienky, a potom s nimi „interaguje“, niektoré z nich po čase odmieta, ďalšie stále prijíma, iné modifikuje. Samozrejme, kľúčová je tu konfrontácia (interakcia) s prostredím. Súčasťou prostredia sa stávajú aj ľudské myšlienky, ktoré uľahčujú a podporujú kognitívne aktivity [1], pozri nižšie. Deje sa to vlastne (takmer) nepretržite. Až vďaka interakcii opísaného typu možno odlíšiť korektné usudzovanie.

Analyzujme môj obľúbený príklad (stroskotalo na ňom veľa študentov, dosť tvrdo trénovaných v korektnom argumentovaní). Vyplýva záver „*Niektoré zvieratá sú hnedé*“ z predpokladov „*Niektoré medvede sú hnedé*“ a „*Niektoré zvieratá sú medvede*“? Predstavíme si dôležitý kognitívny nástroj, ktorý vyvinula interaktívna myseľ, techniku protipríkladu. Uvažujme tým istým spôsobom ako v uvedenom príklade, iba zvieratá nahraďme párnymi číslami, medvede nepárnymi číslami a hnedé prvočíslami. Po tomto nahradení sa dostaneme od dvoch pravdivých predpokladov k nepravdivému záveru. Skonštruovali sme protipríklad, ktorý ukazuje, že usudzovanie opísaným spôsobom nezaručuje, že od pravdivých predpokladov sa dostaneme k pravdivým záverom.

O takomto usudzovaní môžeme povedať, že nie je korektné. A naopak: ak nemožno nájsť k danému úsudku protipríklad, potom je *korektný*. Pri každej možnej interpretácii, v ktorej sú pravdivé predpoklady, je pravdivý aj záver. Zaručene sa dostaneme od pravdivých predpokladov k pravdivému záveru. Zdôraznime, že obrovskú silu získala kognícia práve tým, že začala brať do úvahy všetky možné situácie. Reprezentácia, obohatená o konštrukt všetkých možných situácii, umožňuje dôkladnú argumentáciu, jemné rozlišovanie spoľahlivých a nespoľahlivých argumentov. Tento konštrukt možno považovať za jeden z nástrojov, ktorými – podľa prenikavej Dennettovej [1] myšlienky – ľudia zosilňujú svoje schopnosti (medzi nimi aj schopnosť rozmýšľať) tým, že nechávajú stopy v prostredí.

Rozšírili sme našu observačnú bázu o konštrukt protipríkladu a všetkých možných svetov (stavov, situácií, modelov), ktorý je základom sémantík, charakterizujúcich usudzovanie. Zdôraznime, že tento konštrukt má nesporný kognitívny obsah.

Treba sa zastaviť ešte pri produktívnych a užitočných úsudkoch, ktoré však nezaručujú, že sa dostaneme od pravdivých predpokladov k pravdivým záverom. Ide o hypotetické (nemonotónne) úsudky. Jeden prístup k ich sémantickej charakterizácii je založený na tom, že nepracujeme so všetkými možnými situáciami, ale iba s relevantnými. Ďalší pracuje iba s takými modelmi, ktoré majú vlastnosť stability: dovoľujú akceptovať iba množiny takých hypotéz, ktoré spolu dávajú zmysel, nevyvracajú sa vzájomne (podrobnosti v [9], ide o preferenčné alebo minimálne vyplývanie, stabilné modely, extenzie defaultových teórií, apod.). Iba podotknime, že z výpočtového hľadiska ide o problémy primerane charakterizované až pomocou nedeterministických výpočtových zariadení. Spomínané konštrukty (relevancie a stability) majú tiež nesporný kognitívny obsah a rozširujú observačnú bázu našej teórie o sémantiku umožňujúcu rozlíšiť, kedy je hypotetické usudzovanie produktívne, „čisté“ a kedy nie.

5 Základné intuície

V tejto časti sa budeme venovať neformálnym opisom základných konštrukcií použitých v našej formalizácii.

- *Opis 1 (Schéma)* Rýchle usudzovanie, skákanie k záverom je založené na schémach. Schémy sú zhluky elementárnych presvedčení (belief). Každý usudzujúci agent disponuje nejakou množinou schém.
- *Opis 2 (Udalosť)* Usudzujúci agent pôsobí v nejakom prostredí. V prostredí sa odohrávajú nejaké udalosti (máme namysli aj udalosti vo vnútornom prostredí agenta – zámery, túžby, ciele). Schémy sú aktivované udalosťami. Často sa aktivujú viaceré schémy paralelne.
- *Opis 3 (Agent)* O agentovi nám stačí predpokladať iba toľko, že na nejaké vstupy (podnety) odpovedá nejakými výstupmi (reakciami), a že na toto správanie nepotrebuje vonkajšie riadenie.
- *Opis 4 (Usudzujúci agent)* Usudzujúci agent je agent, ktorý má aspoň nasledujúce ďalšie vlastnosti. Disponuje viacerými schémami. V každom časovom okamihu má aktivovanú nejakú schému, ktorú nazývame aktuálnou. Aktuálna schéma je aktivovaná nejakou udalosťou. Agent je schopný skákať k záverom na základe aktuálnej schémy. Môže schémy aj modifikovať (takto získava nové znalosti).

Modifikuje ich na základe istých udalostí. Ďalšie typy udalostí (signalizujúce isté problémy so skákaním k záverom) môžu spôsobiť, že agent „konzultuje“, „volá“ ideálneho agenta a žiada ho o odpoveď na nejakú otázku (tento agent používa odvodzovanie ako podklad pre odpovedanie). Spomínané volanie môže (nedeterministicky) aktivovať viacerých agentov, pričom každý z nich môže používať iný typ odvodzovania.

- *Opis 5 (Ideálny agent)* Úlohou ideálneho agenta je odpovedať na otázky, ktoré mu kladie usudzujúci agent. Odpovede získava odvodzovaním. Ideálny agent má pri odvodzovaní prístup k báze poznatkov. Je schopný získavať nové poznatky (rozširovať bázu poznatkov). Odvodzuje dôsledky z viet v báze znalostí. Súčasťou tejto schopnosti je vyselektovanie relevantných viet bázy znalostí, o ktoré opiera odvodzovanie. Niektoré z jeho odpovedí môžu znamenať, že nemá dostatočné informácie pre adekvátnu odpoveď. V takom prípade sa usiluje získavať nové informácie alebo signalizuje usudzujúcemu agentovi, aby ich získaval tiež. Niektoré z udalostí môžu znamenať atak proti niektorej časti bázy znalostí. Tieto ataky môže ideálny agent odraziť argumentáciou alebo akceptovať a modifikovať bázu znalostí. Ideálny agent je schopný zdôvodňovať (vysvetľovať) svoje odpovede.

6 Technické detaily

Prvú aproximáciu teórie, založenej na uvedených opisoch a intuíciách predstavuje [11]. Schémy sú reprezentované množinami literálov. Alternatívou k tomuto rozhodnutiu je implementácia schémy čiastočnými funkciami, ktoré dávajú odpovede na istú triedu otázok. Prostredie je reprezentované spoločenstvom agentov. Iba niektorí z agentov sú usudzujúci agenti (alebo ideálni agenti). Udalosti sú implementované v komunikačnom jazyku: každej udalosti v prostredí zodpovedá nejaká správa komunikačného jazyka. Komunikácia v tomto jazyku modeluje dynamiku prostredia. Správanie agentov je modelované výmenou správ. Výmena správ beží podľa istých protokolov. Tieto protokoly však umožňujú nedeterminizmus – v danej situácii možno reagovať podľa rôznych protokolov. Samozrejme, možno si predstaviť oveľa komplikovanejšiu reprezentáciu dynamiky prostredia.

Jednotlivé usudzovacie operácie možno realizovať pomocou spomínaných protokolov. Na otázky sa dá odpovedať buď priamo na základe schémy alebo pomocou sady dostupných argumentácií. Niektoré z udalostí odpaľujú argumentáciu. Pri nej môžu ideálni agenti používať rôzne logiky, rôzne dlhú či komplikovanú argumentáciu. Niektoré z týchto možností poskytujú prostriedky menej náročné na čas, napríklad usudzovanie na základe stabilných modelov predspracovaných a zakódovaných v Kripkeho štruktúrach.

Bázy poznatkov sú reprezentované ako dynamické logické programy. Ide o množiny zovšeobecnených logických programov, na ktorých je definovaná relácia preferencie. Na základe tejto relácie sa riešia prípadné konflikty medzi programami. Táto reprezentácia je dynamická, umožňuje modularitu, revízie a vývojaschopnosť. Kvôli názornosti: Udalosti možno reprezentovať správami ako napríklad (a, b, ASK(q)) alebo (a, b, TELL(S)).[6] Reakciu možno oprieť o rôzne protokoly. Jeden z nich určuje, že treba

[6] Agent a kladie agentovi b otázku q, resp. oznamuje mu množinu viet S.

vyhodnotiť *q* vzhľadom na schému, aktivovanú otázkou *q* a odoslať získanú odpoveď od agenta *b* k agentovi *a*. Iný sa môže pýtať ideálneho agenta, či informácia *S* je zlučiteľná s informáciou, čo má v báze poznatkov, atď.

7 Záver

Projekt je stále v počiatočnom štádiu. Popri domyslení základných konštrukcií ide aj o experimentálne implementácie s využitím štruktúr, ktoré v istom zmysle slova reprezentujú explicitne (všetko) to, čo možno získať (odvodením alebo skákaním k záverom). Pôvodne v [11] sa používali teoretické konštrukty z oblasti dynamického logického programovania, najmä z jeho sémantickej charakterizácie pomocou dynamických Kripkeho štruktúr. Tieto štruktúry majú blízko ku schémam, a preto podporujú dobrú kooperáciu schém a báz poznatkov. Správanie funkcií, aktivujúcich schémy, definujúcich reakcie na správy atď. možno zlepšovať, učiť na základe ohodnocovania dosiahnutých výsledkov. Tu sa črtá spolupráca znalostného prístupu s takými technikami ako je evolučné programovanie. Skákanie k záverom možno postupne vylepšovať (učiniť ho menej chybovým). Množina (paralelne pracujúcich) schém sa zdá byť perspektívnejšou štruktúrou ako jediná aktívna schéma. Dôležitý je mechanizmus aktivácie schém. Ukazuje sa, že je potrebná bohatá trieda schém a pružnosť pri ich aktivovaní, mechanizmus aktivácie schopný učiť sa, optimalizovať svoje správanie.

Veľká pozornosť v súčasnom výskume hypotetického usudzovania sa venuje výpočtovým a implementačným stránkam. Veríme, že predložený prístup môže byť jednou z perspektívnych ciest ako skĺbiť rýchle výpočty s logicky korektnými a dôkladnými výpočtami. Na záver chcem už iba zdôrazniť význam logiky a umelej inteligencie, založenej na logike pre štúdium kognície.

Literatúra

[1] Dennett, D.: *Typy myslí*. Kaligram 1997.
[2] Flach, P.: *An analysis of various forms of 'jumping to conclusions'* In Jantke, K. (ed.), Proc. 3rd Int. Work. on Analogical and Inductive Reasoning, LNAI, vol. 642, 1992.
[3] Fodor, J.: *The mind doesn't work that way*. MIT Press 2000.
[4] Ginsburg, M.: *Readings in Nonmonotonic Reasoning.* Morgan Kaufmann Publ. 1987.
[5] Marek, W.: *NMR - advances, future directions and few questions*. Workshop NMR 2000.
[6] McCarthy, J.: *Epistemological problems of artificial intelligence*. Proc. IJCAI'77, 1038- 1044.
[7] McCarthy, J., Hayes, P.: *Some philosophical problems from the standpoint of artficial intelligence.* In Meltzer, B., Michie, D. (eds.), Machine Intelligence 4. Edinburgh University Press 1969.
[8] Minsky, M.: *A framework for representing knowledge.* In Winston, P. (ed.), The Psychology of Computer Vision. McGraw-Hill 1975.
[9] Šefránek, J.: *Inteligencia ako výpočet*. Bratislava: IRIS 2000.
[10] Šefránek, J.: *Kognícia bez mentálnych procesov*. V: Rybár, J., Beňušková, Ľ., Kvasnička, V. (editori): Kognitívne vedy. Bratislava: Kaligram 2002, 200-256.

[11] Šefránek, J.: *Jumping to conclusions*, 2001, http://www.ii.fmph.uniba.sk/~sefranek
[12] Wiedermann, J.: *Výpočetní meze kognitivních systémů*. V: Kelemen, J., Kvasnička, V. (editori): Kognice a umělý život II. Opava 2002, 253-264.

Hierarchické zobecňující gramatiky

Tomáš Brandejský[1]

Abstrakt. Článek představuje hierarchické zobecňující gramatiky. Jedná se o specielní typ gramatik, které jsou zkonstruovány rozšířením prostoru slova gramatiky, transformací kontextové gramatiky na parametrickou gramatiku a zobecněním této gramatiky a vytvořením hierarchie takovýchto gramatik. Přístup založený na hierarchických zobecňujících gramatikách dovoluje popisovat mnoho systémů z oblasti umělé inteligence a soft-computingu.

1 Úvod

V oblasti umělé inteligence a kognitivních věd známe příklady mnoha modelů lidského myšlení, které „mají cosi společného“, ale přesto jsou popisovány pomocí jiného typu matematiky, analyzovány jinými nástroji, implementovány odlišnými typy algoritmů a proto se zdají být neslučitelné přes určitou podobnost svého chování. Zřejmě i proto vyzněla poněkud naprázdno jednotící snaha L. Zadeha zvaná soft computing. Například algoritmy genetického programování i neuronové sítě jsou schopny se učit z příkladů. Produkční expertní systémy i algoritmy genetického programování dokáží vytvářet struktury, podobně jako některé typy gramatik, buňky v celulárním automatu jsou popsány konečným automatem, který je ekvivalentní gramatice atd.

Gramatiky se v tomto směru zdají být určitým jednotícím motivem, protože o většině výše uvedených systémů je známo, že je dokážeme popsat určitými speciálními typy gramatik.

2 Zobecněný pohled na písmena a slova gramatik

Existuje velké množství systémů, které jsou popsatelné v konečných diskrétních prostorech a modely téměř všech dalších jsme do těchto prostorů schopni transformovat. Nakonec reálná čísla v té podobě, jak jsou reprezentována v našich počítačích rovněž tvoří konečnou a spočitatelnou množinu (podmnožinu množiny všech reálných čísel, která nekonečná skutečně je).

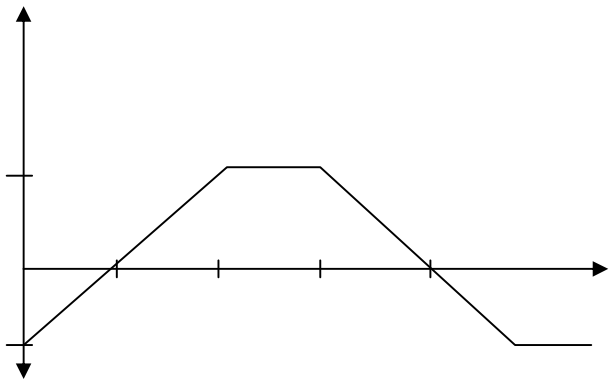
Pohyb systému v prostoru jsme schopni popsat diferenciálními rovnicemi, ale také aproximovat mnoha různými způsoby. Můžeme provést regresní analýzu a systém popsat poněkud jinými rovnicemi, než je řešení diferenciální rovnice, ať již proto, že přesnou diferenciální rovnici neznáme, nebo že její řešení je příliš komplikované. Můžeme také převést diferenciální rovnice na diferenční, na kvalitativní model a pod. Pokud je provedena diskretizace prostoru (resp. jeho granulace jako např. v případě fuzzy popisu), můžeme pohyb systému také popsat gramatikou (viz např.[1]), leckdy regulární.

Například kvalitativní model kmitavého pohybu zachyceného na obr. 1 můžeme také popsat gramatikou, viz obr. 2.

Jiným systémem, který si tak říkajíc žádá o popis pomocí gramatik jsou celulární automaty. Celulární automat je systém skládající se z mnoha navzájem propojených

[1] Fakulta dopravní ČVUT v Praze, Konviktská 20, 110 00 Praha 1, E-mail:Brandejsky@fd.cvut.cz

konečných automatů. Tyto automaty bývají mnohdy regulární a proto popsatelné pomocí regulární gramatiky. Jsou také zpravidla velmi jednoduché, takže jejich stav můžeme reprezentovat jediným písmenem. Nabízí se otázka, proč bychom tak jednoduchý systém, jakým je např. známá hra Life, nemohli popsat jediným slovem a jedinou gramatikou. Jediným důvodem, který tomu brání, je reprezentace slova pomocí lineárního řetězce písmen, zatímco pro reprezentaci celulárního automatu je zapotřebí dvou- či vícerozměrné pole.



Obrázek 1. Kvalitativní průběh kmitavého pohybu.

-- -> 0-
0- ->+0
+0 ->++
++ -> 0+
0+ ->-0
-0 -> --
Abeceda: +, 0, -

Obrázek 2. Gramatika popisující kvalitativní průběh kmitavého pohybu

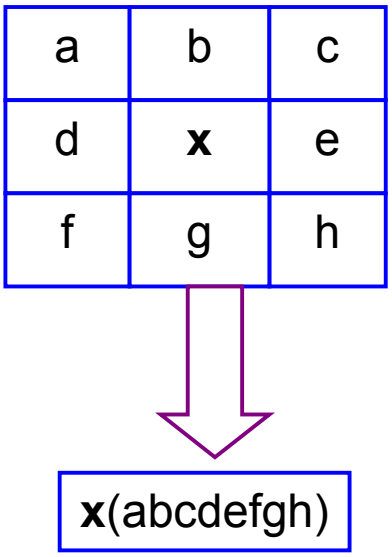
2.1 Gramatika s vícerozměrným slovem

Nicméně lineární uspořádání písmen ve slově gramatiky je poplatné pouze lidské řeči, pro jejíž výzkum Chomsky gramatiky použil. Tomu u kontextových gramatik odpovídá i existence pouze levého a pravého kontextu. Žádný další důvod pro toto omezení není, proto nadále budeme předpokládat, že slovo gramatiky je n-rozměrný soubor písmen, například hyperkrychle. Dimenze n se v průběhu výpočtu v dané gramatice nemění (proměnná dimenze by přinášela problémy s tvorbou souboru pravidel a vedla by ke vzniku gramatik vyššího řádu). Nicméně výrazně narůstá počet kontextů. Pokud bychom uvažovali pouze osové směry, i tak těchto kontextů bude $2n$! Například pokud bude slovo dvourozměrné (matice písmen), vedle levého a pravého kontextu získáme i horní a dolní kontext.

Protože prostor slov gramatiky je konečný, můžeme tuto multidimenzionální gramatiku převést na jednorozměrnou zvláštní mapovací funkci. Tato funkce přináší lokalitu do globálně platných pravidel. Z čistě pragmatického hlediska je výhodné použít

mapovací funkci, která lokální kontext převede na parametr písmena, ke kterému se tento kontext vztahuje, viz. Obr. 3.

Výše načrtnutý koncept můžeme dále zobecnit. Především parametr nemusí reprezentovat všechny potenciální prvky slova, ale může je libovolně vybírat a přeskupovat. Konverze z prostoru slova do prostoru parametru je z obecného hlediska funkce nejen slova a polohy daného písmena v něm, ale i času. Prostor parametru má obecně jinou (nižší) kardinalitu, než celé slovo.



Obrázek 3. Znázornění převodu vícerozměrného kontextu na parametr.

S jistými omezeními bychom mohli nahradit mapovací funkci gramatikou. Tato gramatika ovšem v principu může mít opět vlastnosti výše uvedeného rozšíření.

2.2 Zobecnění pojmu písmena gramatiky

Písmeno gramatiky nemusí nutně představovat znak z ASCII tabulky, ale může reprezentovat jakýkoli jednoznačně identifikovatelný algebraický objekt – např. vektor hodnot, nebo rámec. Písmeno by mohlo reprezentovat i celou gramatiku. Těchto skutečností v dalších pasážích tohoto článku využijeme pro modelování mnoha různých systémů.

V dalším textu zpravidla nebudeme uvažovat gramatiky v Chomského smyslu, ale L-gramatiky (viz [2]). L-gramatiku můžeme totiž mnohdy chápat jako (rekurzivní) diskrétní funkci pracující na prostoru stavového slova (injektivní zobrazení). Pokud by L-gramatika byla stochastická, mohli bychom hovořit o korespondenci.

2.3 L-gramatika s parametrickým slovem

V dalším textu budeme uvažovat rozšíření klasické parametrické L-gramatiky ve smyslu chápání parametru. Běžně je parametr chápán jako jeden prvek z množiny možných parametrů a tento parametr se nevyskytuje ve vlastním derivačním pravidle. Pokud by se

parametr vyskytoval v podmínce pravidla, pak by mohl buď nahrazovat kontext (o což se v dalším textu pokusíme), nebo by mohl nahrazovat tabulkovou gramatiku.

Tabulková gramatika je gramatika s několika alternativními soubory pravidel, které jsou přepínány zpravidla z vnějšku gramatiky. Tabulkové L-gramatiky bývají využívány např. k modelování jednotlivých růstových fází rostliny.

Pokud by se parametry písmen gramatiky vyskytly v podmínkách pravidel, byl by tento soubor pravidel převoditelný na tabulkovou gramatiku.

Důvody, proč nás tato varianta zápisu tabulkové gramatiky zajímá jsou následující:

- parametr nemusí být jednotný pro všechna písmena slova gramatiky. Pokud by gramatika byla např. kontextová, jednotlivých variant tabulkové gramatiky by bylo vzhledem ke kombinatorické explozi velké množství. Konkrétní variantu tabulkové gramatiky by totiž nepředstavoval jeden univerzální parametr, ale vektor jehož kardinalita by odpovídala kardinalitě slova gramatiky. Počet souborů pravidel ekvivalentní tabulkové gramatiky by byl roven kardinalitě množiny přípustných hodnot parametrů umocněné kardinalitou slova gramatiky.
- parametrem v principu nemusí být jediný znak, ale i soubor znaků (slovo). Tato varianta opět přináší nepodstatnou změnu, neboť množina slov je opět ekvivalentní množině symbolů (reprezentujících jednotlivá slova), ale umožní nám konstrukci hierarchické zobecňující gramatiky.

Pokud ale bude parametrem gramatiky slovo, změní se způsob vyhodnocování takového parametru, zvláště když připustíme výskyt parametru na levé straně derivačních pravidel.

Parametrické slovo totiž nemusí být zkoumáno na shodu s podmínkou pravidla jako celek, ale nabízejí se i další možnosti:

substringem parametrického slova je pattern uvedený v podmínce pravidla

parametrické slovo obsahuje všechny znaky uvedené v podmínce pravidla

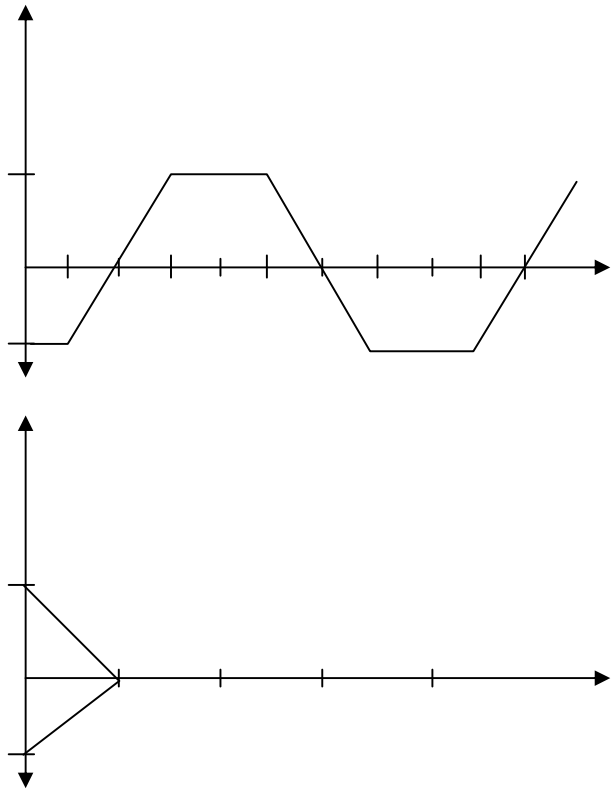
Pokud by byla užita tato vyhodnocovací pravidla, počet variant ekvivalentní tabulkové gramatiky zřejmě opět o několik řádů vzroste a výše naznačený zápis představuje mnohem informačně koncentrovanější formu.

Jinou variantou interpretace parametrické gramatiky, kde jsou hodnoty parametrů uvažovány i v předpokladech pravidel je uvažování písmena s konkrétní hodnotou parametru jako zvláštního písmena abecedy s tím, že současně původní písmeno bez parametru může zastupovat všechna písmena odpovídající tomuto písmeny s jednotlivými možnými parametry.

3 Obecnější pohled na mapovací funkci

Mapovací funkce je nazvána dle svého základního úkolu – mapovat prvky jednorozměrného slova parametru do n-rozměrného slova původní gramatiky. Protože tato funkce může být v čase proměnná, může docházet k deformacím prostoru, nad kterým původní gramatika pracuje. Následkem toho je tato gramatika ovlivněna, generuje jiná slova, aniž by bylo nutno změnit soubor pravidel této původní gramatiky. Tato změna souboru pravidel (původní) gramatiky by mohla přinášet komplikace se zajištěním konzistence nového souboru pravidel. Nicméně tyto dva přístupy (změna pravidel gramatiky v závislosti na aktuálním slovu a deformace prostoru – kontextu – pomocí změny mapovací funkce) jsou navzájem ekvivalentní. Např. pokud bychom změnili kontext gramatiky

z obr. 1 a 2 tak, že by stavové slovo netvořil minulý a současný stav, ale předminulý a současný, nebo současný a současný, dostali bychom průběhy zachycené na obr. 4:

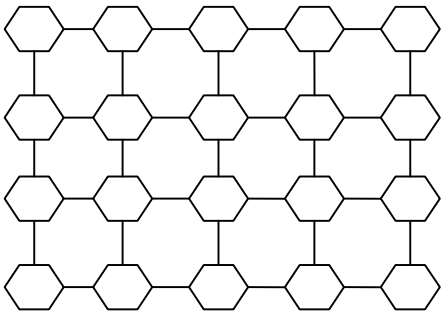


Obrázek 4. Průběhy generované gramatikou popsanou v Obrázku 2 při změně kontextu.

Deformace geometrie prostoru se mnohdy využívá v algebře pro zjednodušení výpočtu. Jako příklad je možno uvést logaritmování.

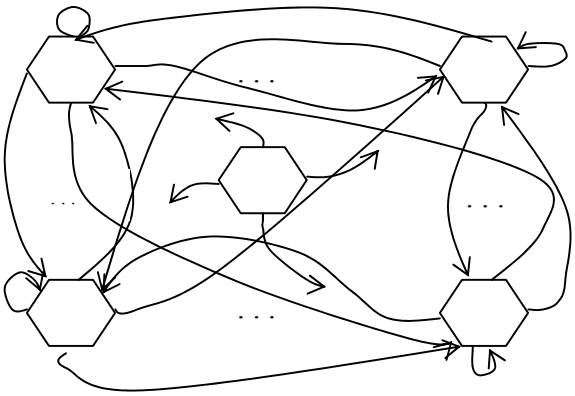
Další příklady transformací si můžeme ukázat na známé počítačové simulaci života kolonie buněk – programu LIFE.

Pokud by systém byl popsán gramatikou s vícerozměrným slovem (viz kap. 2.1), pak by stav každé buňky popisovalo právě jedno písmeno slova gramatiky. Budoucí stav dané buňky (daného znaku ve slově) závisí i na okolních buňkách. Tento odkaz na okolní buňky můžeme pomocí mapovací funkce nahradit parametrickým slovem každého jednotlivého písmena. Pokud je stavové slovo tvořeno sousedními písmeny, získáme standardní celulární automat hry LIFE, viz obr. 5:



Obrázek 5. Schéma celulárního automatu hry LIFE.

Pokud by ale například všechny vazby vedly do neměnných bodů sítě, získáme jinou zajímavou topologii, viz obr. 6, kde stav všech bodů libovolně velké sítě je určen vedle jejich předchozího stavou pouze stavy vrcholových (rohových) buněk v této matici:



Obrázek 6. Příklad složitější topologie, kdy se již vazby mezi buňkami stávají orientovanými.

U složitějších topologií se stávají vazby mezi buňkami automatu (písmeny slova gramatiky tvořícími mnohorozměrný kontext) orientovanými. Rovněž je mnohdy třeba řešit otázku které buňky vlastně tvoří kontexty okrajových buněk (pokud je slovem *n*-rozměrná matice písmen). Vedle obvyklého přístupu že první písmeno ve slově (vektoru písmen) nemá levý kontext atd. si lze představit i jiná pojetí – kontextem levých písmen jsou písmena pravá atd., což vede de facto k *n*-dimenzionální kouli. Příklad zachycený na obr. 6 zase vychází z pojetí pevně daných bodů tvořících kontext bez ohledu na polohu konkrétního písmene ve slově – k zanedbání vlivu lokality.

Nejen z implementačního hlediska je výhodné mapovací funkci resp. gramatiku rozdělit na transformační funkci a mapovací gramatiku. Transformační funkce slouží k vlastní transformaci slov z abecedy původní gramatiky do abecedy nadřazené mapovací gramatiky. To znamená, že je využita k výběru písmen z původního slova, převodu písmen na jejich relativní adresy, resp. objevování libovolných relací mezi písmeny původního slova; může být i využita k redukci rámců představujících jednotlivá písmena (vynechání irelevantních atributů) a pod. To dovoluje popsat v mnoha případech přidruženou mapovací gramatiku jako nereducibilní, což přináší mnohé, nejen teoretické výhody.

V tomto systému budeme původní gramatiku považovat za podřízenou (neb její výstupy závisí na výstupech mapovací gramatiky a mapovací gramatiku budeme považovat za nadřazenou.

Uvedené argumenty ale současně znamenají nejen že dvojice transformační funkce – mapovací gramatika představuje morfismus mezi abecedou původní a nadřazené gramatiky resp. mezi jejich slovy. Můžeme zde také formulovat tvrzení, že těchto dvojic (transformační funkce a mapovací gramatika) můžeme zkonstruovat sice konečné množství, ale zpravidla vyšší než jedna. Tato mnohoznačnost je způsobena redukcí informace v transformační funkci (a případně i v mapovací gramatice).

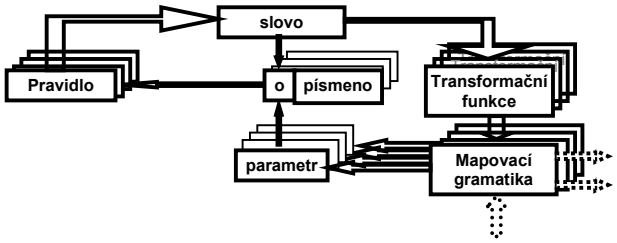
Mimo to mohou transformační funkce a mapovací gramatika výrazně ovlivňovat vlastnosti podřízené gramatiky.

Mapovací gramatika může generovat pouze odkazy v rámci dané množiny slov, pak tedy má vlastnosti faktorové grupy. (Pokud by např. odkazy mířící ven ze sítě LIFE vstupovaly opět na opačné straně, protože kardinalita slova nadřazené gramatiky je konstantní). Opačnými příklady nemajícími tyto vlastnosti jsou buď nekorektní gramatiky, nebo gramatiky do nichž vstupují externí signály, na které se mohou odkazovat přepisovací pravidla, viz kapitola 5.8.

Jiným řešením jsou specielní pravidla generující prázdné odkazy a zvláštní symbol v abecedě vazeb reprezentující tento odkaz, který by byl nevytvořitelný transformační funkcí.

4 Definice hierarchické zobecňující gramatiky

Navržené rozšíření představy gramatik je z hlediska datových toků zachyceno na obr. 7:



Obrázek 7. Datové toky v hierarchické zobecňující gramatice.

Z tohoto obrázku je zřejmé, že vedle klasické výpočetní smyčky gramatiky (slovo → rozklad na písmena a přiřazení parametru → pravidla → slovo) zde probíhá i druhá smyčka (slovo → transformační funkce → mapovací gramatika → parametry písmen → pravidla → slovo s přiřazeným parametrem), která prostřednictvím parametrů písmen zavádí do výpočtu zpětnou vazbu. Současně zde dochází k určitému porušení lokality, narušení sémantické konzistence a pokud je mapovací gramatika opět hierarchickou zobecňující gramatikou, pak i ke zvláštnímu běhu času. Tyto skutečnosti připomínají propletené hierarchie, které Hofstadter pokládal za nutnou podmínku inteligentního chování ve své knize [3] (možnosti využití gramatik pro modelování kreativity nepřímo věnoval knihu [4]). Následující odstavce budou věnovány rozboru důsledků těchto tří vlastností.

4.1 Porušení lokality

Bezkontextová gramatika pracuje čistě lokálně. Na dané písmeno ve slově je aplikováno určité pravidlo a nedochází ani k ovlivnění okolními písmeny, ani nedojde k jejich změně, nebo k ovlivnění výběru pravidel, která budou na ně aplikována. Kontextová gramatika přináší ovlivňování okolím do určité vzdálenosti dané mohutností levého a pravého kontextu. Nicméně i tento kontext je vztažen k danému bodu a je tedy do jisté míry lokální. Hierarchická zobecňující gramatika přináší možnost selektivního výběru písmen kontextu, jejichž relativní poloha je sice vztahována k danému písmenu, ale není neměnná a může nabývat hodnot kteréhokoli písmene slova gramatiky. Z tohoto hlediska tvoří transformační funkce (spolu s nadřazenou mapovací gramatikou a převodem neparametrické gramatiky na parametrickou) faktorovou grupu nad prostorem slova gramatiky.

Narušení lokality vedle již zmíněné částečné ekvivalence se změnou pravidel dovoluje vytvářet relace mezi vzdálenějšími body, než je dáno jejich pořadím ve slově. Obdobně pracují mnohé genetické algoritmy (změna pořadí genů, resp. jejich volba pro operace křížení, v závislosti na jejich ohodnocení) a k narušování lokality dochází i např. u algoritmů analogického a metaforického uvažování.

4.2 Narušení sémantické konzistence

K narušení sémantické konzistence může v hierarchické zobecňující gramatice dojít několika způsoby.

V první řadě písmena ve slově gramatiky mohou být v počátečním stavu uvažována jako bezparametrická. Před aplikací pravidel je písmenům jejich parametr doplňován z nadřazené mapovací gramatiky (resp. je o hodnotu získanou z nadřazené gramatiky doplňován). Pravidla jsou v tomto směru asymetrická, neboť parametr reprezentující zobecněný kontext se v nich vyskytuje pouze na levé straně. Pokud by gramatika byla sama o sobě parametrická, pak by před aplikací pravidel docházelo k rozšíření prostoru parametrů, resp. na levé straně pravidel by vystupovala písmena se dvěma parametry.

Ke druhému porušení sémantické konzistence dochází vlastním převodem kontextu na parametr. Transformační funkce není jen funkcí kontextu, ale také času. To znamená, že jí produkované hodnoty mohou být nezávislé na původním slově.

4.3 Zvláštní běh času

Do jednoho taktu (cyklu) gramatiky jsou rekurzivně vkládány cykly nadřazených gramatik. To znamená, že s čím vyšší abstraktní úrovní daná gramatika pracuje, tím jemnější dělení času používá. Vše dále záleží na způsobu práce konkrétní gramatiky. Nadřazené gramatiky totiž mohou vykonat právě jeden cyklus (pak vystupují jako jednoduché funkce), více cyklů do splnění dané podmínky, nebo neomezený počet cyklů až do zastavení (což není aplikovatelné na gramatiky s nekonečnou dobou výpočtu, tedy na gramatiky s cykly v derivačním stromu a na mnohé gramatiky s rostoucí délkou slova).

4.4 Implementace mnohoznačnosti

Jak již bylo řečeno v kap. 3, práce s abstrakcí vede k mnohonásobným interpretacím reprezentujícím dílčí pohledy na strukturu a funkce systému. Pokud mapovací gramatika a případně i transformační funkce generují mnohonásobná zobecnění původního systému (slova gramatiky), pak můžeme buď z nich vybrat jedinou, nebo řešit všechny (eventuelně vybrat některé).

Podkladem pro volbu jediné abstrakce může být rozhodnutí o její přijatelnosti z nadřazené gramatiky, zavedení ohodnocovací funkce, nebo eliminace neurčitosti daná zavedením pravidla platnosti první interpretace.

Pokud budeme uvažovat variantu užití více (všech) zkonstruovatelných zobecnění, pak zřejmě jen sporadicky bude užívána varianta, kdy by parametr předávaný písmenům původní gramatiky tvořilo sjednocení všech interpretací (i když i to může být výhodné v konstruktivních úlohách ve spojení s chápáním parametru uvedeného v pravidle jako vzoru, který je substringem skutečného takto vypočteného parametru). Výpočetně zvládnutelnější se jeví možnost generovat a poté paralelně vyhodnocovat a dále rozvíjet více interpretací. I při použití této techniky je třeba postupovat velmi opatrně vzhledem ke kombinatorické explozi počtu variant. Je pak vhodné (je-li to možné) gramatický systém doplnit o mechanismy hashování a vylučování shodných větví vývoje, např. o mechanismus větví a mezí a pod.

4.5 Problém gramatiky s proměnnou délkou slova

Pokud v hierarchické zobecňující gramatice dochází ke změně délky slova v podřízené gramatice, odráží se tato skutečnost v nadřazené mapovací gramatice. Dochází k tomu proto, že písmena nadřazené gramatiky představují adresy písmen v původním slově. Zkrácení slova v podřízené gramatice pak některá písmena nadřazené gramatiky (adresy) činí irelevantními. Rozšíření slova v podřízené gramatice naopak vyvolává potřebu přidání písmene do gramatiky nadřazené.

Pokud je z gramatiky odebráno písmeno, je možno zrušit i s ním spojená pravidla, především pak ta, ve kterých se vyskytuje na pravé straně.

Pokud je do gramatiky přidáno písmeno, vzniká problém, že toto písmeno se nevyskytuje ani v předpokladu, ani v důsledku žádného pravidla. Řešením tohoto problému může být automatické vygenerování pravidel dle vzoru (jiného písmena), pokud mají všechna písmena abecedy analogické vlastnosti, nebo pokud jsme schopni nějaké písmeno explicitně označit jako vzor pro dané nové písmeno.

Jiná situace nastává při rozšíření/zúžení slova nadřazené gramatiky (rozšíření slova parametru podřízené gramatiky). Změna délky slova parametru si vynucuje jiné mechanismy práce s tímto slovem.

Pokud byla délka slova parametru konstantní, byla množina slov parametru spočetná a tedy platily závěry kapitoly 2.3. Při proměnné délce slova je problematické uvádět toto slovo v předpokladu pravidla jako celek. Jako mnohem užitečnější (a duchu gramatiky bližší) se jeví otázka zda slovo parametru obsahuje určitý vzor (např. substring). Takovéto řešení by bylo i implementačně schůdné.

5 Příklady aplikací

5.1 Kvalitativní simulace

Příklad kvalitativních modelů byl již použit v kapitolách 2 a 3. Kvalitativní model bývá popsán souborem omezujících podmínek, které jsou obvykle zadány tabulkou pravdivostních hodnot. K řešení bývají používány algoritmy řešení soustav omezujících podmínek. Zřejmě je možné tyto algoritmy modifikovat tak, aby soustavu omezujících podmínek převedly na generující gramatiku. V jednoduchých případech je tuto práci možno vykonat i manuálně, jak tomu ostatně bylo i v kap. 2.

5.2 Celulární automaty

Jak již bylo ukázáno v kap. 2.1, hierarchická zobecňující gramatika s jedinou úrovní (mapovací gramatika je standardní gramatika, v tomto případě zřejmě může být i regulární) dokáže modelovat celulární automaty včetně automatů měnících organizaci buněk.

5.3 Umělé neuronové sítě

K modelování umělých neuronových sítí zřejmě opět postačuje jednoúrovňová hierarchická zobecňující gramatika, jejímž základem ale je parametrická gramatika, viz kap. 4.2. Je to dáno nutností v obecném případě mapovat propojení neuronů a jejich váhy (u některých homogenních sítí, jako je např. Hopfieldova síť, tato potřeba odpadá). Vlastní propojení je kódováno v mapovací gramatice, váhy vazeb pak tvoří parametrický vektor daného písmene – neuronu. Gramatika pak může modelovat i vlastní učící algoritmy.

Protože vstupní informace je předávána jako slovo gramatiky na nejnižší hierarchické úrovni, vyvstává před námi otázka zachování informace o váhách a propojeních dané neuronové sítě.

Možným řešením je zachování informace ve vyšších úrovních hierarchie gramatik, tedy v parametrech písmen, tedy transformační funkce pouze opakuje předchozí slovo nadřazené gramatiky. Jinými slovy, při zápisu nového slova gramatiky zůstávají předchozí parametry nezměněny. To znamená, že parametry musejí mít stejný význam pro všechna písmena gramatiky. Shodný obrat bude muset být použit i v následujících modelech.

5.4 Produkční expertní systémy

O produkčních expertních systémech jako OPS/5 nebo CLIPS je obecně známo, že je můžeme považovat za zvláštní variantu gramatik. Přidání nadřazené hierarchie gramatik nám pak dovoluje popisovat obecné hierarchie takových systémů.

5.5 Algoritmy genetického programování a genetické a evoluční algoritmy

Modelování algoritmů genetického programování (GPAs) pomocí gramatik je možné dvěma způsoby. Buď mapujeme stromovou strukturu genu do lineárního slova s proměnnou délkou (hloubka stromu reprezentujícího vytvářenou strukturu bývá proměnná), nebo je zvolena maximální přípustná délka genu a o použití či nepoužití daného písmene rozhoduje sémantická analýza vytvořené struktury. Druhá varianta je pro

modelování pomocí hierarchických zobecňujících gramatik schůdnější. Pokud prohlásíme vektory písmen za jednotlivé geny, pak celý genom představuje 2D slovo. Jednotlivá pravidla gramatiky popisují operace mutace, křížení a ohodnocení a nadřazená mapovací gramatika rozhoduje o pořadí genů a volbě konkrétní operace. Tento způsob modelování chování genetických algoritmů pomocí gramatik byl publikován v práci [5].

5.6 Konceptuální design

Pokud chceme modelovat lidské myšlenkové a rozhodovací procesy při vytváření komplexních struktur, musíme být schopni modelovat pohyb v komplexním prostoru, jehož složitost tak, jak postupuje řešení, narůstá či klesá. Vytvářený systém můžeme modelovat z mnoha hledisek (vytvářet sub-modely, sub-struktury). Současně zařazení každého prvku do modelu přináší další omezující podmínky dané např. pracovními podmínkami přidaného prvku (viz např. [6]).

Hierarchické zobecňující gramatiky mnohé požadavky na prostředí systému konceptuálního designu splňují. Konkrétně jsou hierarchické a provádějí zobecňování popsané struktury (dovolují popisovat systém na více úrovních abstrakce, z různých hledisek), dovolují jak vytvářet strukturu na hierarchicky nejnižší úrovni a zobecňovat ji (vysvětlovat, ohodnocovat), tak i z obecného popisu generovat popis detailnější přes mechanismy ovlivňování parametrů písmen podřízené gramatiky (design zdola nahoru a shora dolů). Gramatika může pracovat i s proměnnou délkou slova, tedy s proměnnou složitostí systému tak, jak je v průběhu konstrukčního procesu zapotřebí.

5.7 Uvažování založené na asociacích a metaforách

Asociace a metafory jsou založeny na vytváření abstraktních modelů a hledání podobností mezi těmito modely a konkrétními situacemi (viz např. [7]). V případě metafor může být těchto úrovní zobecnění i více, neboť i podobnostní míry jsou asociovány konkrétní situací.

Pokud písmena parametrické gramatiky představují rámce (viz kap. 2.2), pak hierarchická zobecňující gramatika je schopna modelovat asociativní síť a zřejmě bude schopna modelovat i pokročilejší metody asociativního a metaforického uvažování, což ale bude ještě třeba potvrdit a prokázat v dalším výzkumu.

5.8 Vztah k numerickým metodám

Protože gramatika generuje z počátečního slova posloupnost nových slov, je jejím zvláštním případem rekurzivní diskrétní funkce. Protože časové okamžiky nadřazené gramatiky jsou vkládány do kroku gramatiky podřízené, určeného intervalem mezi časy t a t+1, byly pro časové okamžiky nadřazené gramatiky zvoleny fiktivní časové okamžiky t+0.3 a t+0.7. Okamžiky t+0.3 a t+0.7 nemají přesný fyzikální význam, pouze označují dva rozlišitelné časové okamžiky s pevným pořadím, které následují po okamžiku t, ale předcházejí okamžiku t+1.

Písmena gramatiky označíme $\{p_i(S_i^{\prime})\}$.

Slovo gramatiky je S.

Funkce představující gramatiku

$$S(t+1) = F(S(t)) \tag{1}$$

nebo pro jednotlivá písmena regulární gramatiky

$$p_i\left(S_i^{\backslash}(t+0.3), t+1\right) = f\left(p_i\left(S_i^{\backslash}(t+0.3), t+0.7\right)\right) \tag{2}$$

Parametr *i*-tého písmena (slovo nadřazené gramatiky) značíme $S_i^{\backslash}$.

Písmena nadřazené (mapovací) gramatiky označujeme

$$\left\{p_{i,j}^{\backslash}\right\}, \text{ resp. } \left\{p_{i,j}^{\backslash}\left(S_{i,j}^{\backslash\backslash}\right)\right\} \tag{3}$$

Transformační funkce z prostoru slov gramatiky do slov nadřazené gramatiky je

$$S_i^{\backslash} = G(S, i) \tag{4}$$

Funkce reprezentující nadřazenou mapovací gramatiku je

$$S_i^{\backslash}(t+0.3) = F^{\backslash}\left(S_i^{\backslash}(t)\right) \tag{5}$$

Zpětná transformace vytvoří parametrickou gramatiku

$$S(t+0.7) = z\left(S(t), \left\{S_i^{\backslash}(t+0.3)\right\}\right) \tag{6}$$

Výsledná funkce pro jedinou úroveň abstrakce má tvar

$$S(t+1) = F(S(t+0.7)) \tag{7}$$

Pokud by gramatika byla definována nad slovy o jednom písmeni, pak by i nadřazená gramatika měla slova o nejvýše jenom písmeni a výslednou funkci pro jednu úroveň abstrakce můžeme napsat ve tvaru

$$S(t+1) = F\left(z\left(S(t), F^{\backslash}(G(S(t), i))\right)\right) \tag{8}$$

Pokud v tomto vztahu položíme

$$F(A, B) = A + c_1 B \tag{9}$$

a

$$G(x, i) = x \tag{10}$$

pak získáváme vztah popisující Eulerovu metodu numerické integrace, kde konstanta c_1 představuje integrační časovou konstantu.

6 Shrnutí

Představované hierarchické zobecňující gramatiky představují poměrně univerzální nástroj pro modelování libovolných informačních systémů. Byly vytvořeny s cílem sjednocení mnoha známých technik, obzvláště z oblasti soft-computingu a umělé inteligence. Současně před námi staví zajímavé nové teoretické otázky.

Například pokud platí, že gramatika 0. typu dle Chomského hierarchie je ekvivalentní Turingovu stroji, pak buď jakákoli hierarchická zobecňující gramatika je redukovatelná na jednoduchou gramatiku nejvýše 0. typu, nebo můžeme zkonstruovat takovou hierarchickou zobecňující gramatiku, která nebude řešitelná na Turingově stroji.

Tyto gramatiky přinášejí i množství dalších otázek, jejichž zodpovězení si zřejmě vyžádá jistý čas, ale slibné vyhlídky nám jej jistě vynahradí.

Literatura

[1] Kolář, J., Štěpánková, O. a Chytil, M.: *Logika, algebry a grafy*. Praha: SNTL 1989.
[2] Lindenmayer A.: Mathematical models for cellular interaction in development I,II. *Journal of Theoretical Biology* **18** (1968) 280-315.
[3] Hofstadter, D. R.: *Goedel, Escher, Bach: an Eternal Golden Braid*. New York, NY: Basic Books 1979.
[4] Hofstadter, D. R. and the Fluid Analogies Research group: *Fluid Concepts and Creative Analogies: computer models of the fundamental mechanisms of thought*. New York, NY: Basic Books 1995.
[5] Brandejský, T.: The Grammatical Model of Genetic Algorithm. In R. Matousek, P. Osmera (eds): *Proc. 12th International Conference on Soft Computing - Mendel 2006*. Brno: VUT Brno 2006.
[6] Chase, S. C.: A model for user interaction in grammar-based design systems. *Automation in construction* **11** (2002) 161–172.
[7] Gentner, D., Holyoak, K. J., Kokinov, B. N., (eds.): *The Analogical Mind: Perspectives from cognitive science*. Cambridge, MA: MIT Press 2001.

Proč je hranice mezi virtuální bytostí a jejím světem neostrá: representace ISMA a sdílené intence

Cyril Brom[1], Jiří Lukavský[2]

Abstrakt. Projekt IVE – inteligentní virtuální prostředí – vyvinutý na MFF UK je simulátor rozsáhlého virtuálního světa obydleného virtuálními bytostmi. Jádrem prostředí je původní model ISMA („intention – suitability – materialisation – advice"), který je novou technikou pro representaci virtuálního světa. Model staví na Gibsonově teorii afordancí a Bratmanově teorii praktického rozhodování. Postupně se ukázalo, že některé výsledky projektu mají i psychologickou a filosofickou interpretaci. V článku nejprve představíme projekt IVE a model ISMA a poté se budeme zabývat tímto „netechnickým" aspektem projektu.

1 Úvod

Virtuální bytostí se většinou rozumí umělý organismus žijící ve virtuálním prostředí, který splňuje následující vlastnosti:

- Je *vtělený* ("embodied") – to znamená interaguje s virtuálním světem pouze prostřednictvím svého těla, na něž působí zákony virtuálního světa.
- Je *komplexní* – to znamená, že vyjma pohybu vykonává několik dalších úkolů, z nichž obvykle alespoň dva jsou konfliktní.

V anglicky mluvících zemích se pro virtuální bytost vžil název *animat* [20]. Animatem se původně většinou rozuměl biologicky plausibilní behaviorální model živého organismu konstruovaný „zdola–nahoru". Poslední dobou je ale termín zobecňován a začíná se používat i například pro protivníky z počítačových her, u nichž plausibilita není klíčová [9]. Při konstrukci se pak často prosazuje i přístup „shora-dolů". Obecný přehled o virtuálních bytostech podává [6].

Virtuální bytosti jsou většinou zkoumány pouze v kontextu jednotlivých aplikací, například právě v počítačových hrách, nebo v malých virtuálních prostředích v kontextu konkrétního problému. Simulace velikých prostředí obydlených různými typy komplexních bytostí zatím stály na okraji zájmu.

V roce 2005 byl na MFF UK vyvinut softwarový projekt IVE – inteligentní virtuální prostředí – jehož hlavním cílem bylo tuto mezeru vyplnit. Jedná se o nástroj pro vytváření simulací rozsáhlých virtuálních světů obydlených virtuálními bytostmi. Dílčí cíle projektu byly následující:

Automatické zjednodušování simulace. S ohledem na rozsáhlost simulace bylo cílem použít techniku level-of-detail pro řídící algoritmus virtuálních bytostí. Touto technikou se myslí automatické zjednodušování simulace v místech, kde se v daný okamžik „nic

[1] Karlova Univerzita v Praze, Matematicko-fyzikální fakulta, Malostranské nám. 2/25, Praha 1, Česká Republika, E-mail: brom@ksvi.mff.cuni.cz
[2] Psychologický ústav Akademie věd ČR, Husova 4, Praha 1, Česká Republika, E-mail: lukavsky@praha.psu.cas.cz

podstatného neděje“. Narozdíl od jejího běžného používání v počítačové grafice však měla být aplikována pro „umělou inteligenci“.

Snadná rozšiřitelnost. Druhým cílem bylo přijmout takovou representaci virtuálního světa, aby při přidání nového objektu do virtuálního světa „za běhu“ simulace mohly virtuální bytosti automaticky s tímto objektem začít interagovat, aniž by bylo nutné manuálně modifikovat řídící struktury virtuálních bytostí a zároveň aniž by se muselo na jejich úrovni použít technik strojového učení. Tato vlastnost významně ulehčuje tvorbu virtuálního světa.

Aplikace je plně funkční a je k disposici pod BSD licencí implementovaná v Javě [11]. Obou dílčích cílů bylo dosaženo a výsledky byly publikovány v [3] a [14]. Ukázkový virtuální svět obsahuje stovku virtuálních lidí (převážně horníků), několik dolů, restaurací a obytných budov. V současné době se pracuje na editoru pro virtuální světy a popis chování virtuálních bytostí, na episodické paměti pro virtuální bytosti [13] a na „drama manageru“, to jest komponentě pro řízení virtuálních bytostí podle předdefinované osnovy příběhu [8].

Aplikace byla původně zamýšlena jako čistě technický nástroj. Postupně se ale ukázalo, že některé výsledky mají i psychologickou a filosofickou interpretaci. Konkrétně se ukázalo, že IVE vytváří most mezi doménou virtuálních lidí (chápanou jako technická disciplína) a ekologickou percepční psychologií, což umožňuje až metafyzické interpretace IVE. Mimovolně jsme také opustili zažité chápání pracující s rozhraním organismus–svět, což může vést až k obecnému obratu v nazírání na to, jak přirozený svět (na úrovni abstrakce, se kterou IVE pracuje) popisovat.

V tomto článku se budeme zabývat druhým z těchto „netechnických“ témat; problémem rozhraní organismus–svět. Diskusi je vždy třeba chápat v kontextu reálně existující aplikace. Výklad začneme popisem úrovně abstrakce přijaté v IVE. Poté popíšeme representaci virtuálních světů v IVE. Následně zavedeme pojem virtuální svět a budeme zpětně diskutovat, k čemu z přirozeného reálného světa se vlastně prvky representace vztahují. Ukážeme, že na přijaté úrovni abstrakce došlo k rozmazání rozhraní organismus–svět a že došlo ke vzniku hranice jiné – mezi statickou a dynamickou složkou světa. Dále budeme zkoumat význam pojmu „sdílená intence“.[3]

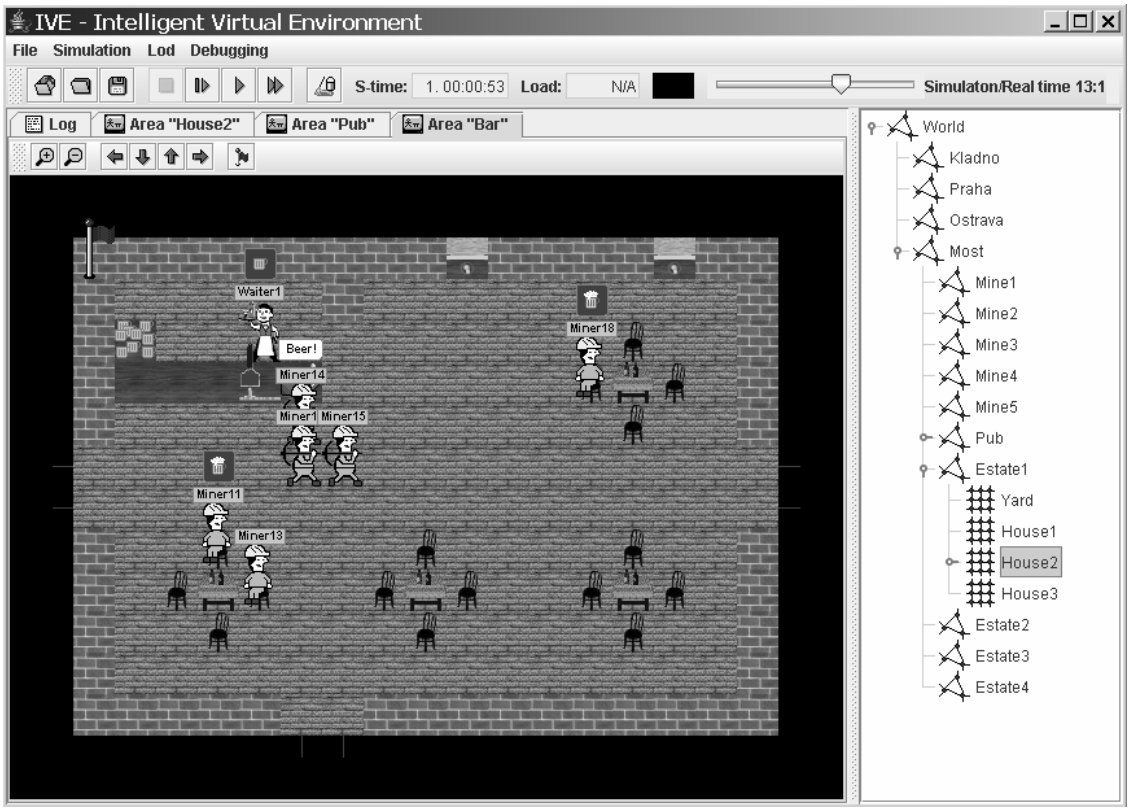
2 IVE

IVE je určeno pro modelování episod a krátkých dějů. IVE je určeno především pro simulace lidských prostředí, proto budeme v dalším textu často nahrazovat obecný pojem „virtuální bytost“ za konkrétnější „virtuální člověk“ (i když IVE samozřejmě umožňuje modelovat, dejme tomu, virtuálního psa). IVE vychází z abstrakce na úrovni objektů a akcí trvajících několik málo vteřin. IVE nepracuje s učením, ani s emocemi. Nyní tyto aspekty simulace rozvedeme podrobněji.

1. Úroveň episod a krátkých příběhů. Aplikace je stavěna pro simulace světa tak, jak ho vidíme, jak se s ním běžně setkáváme. Přesněji je zamýšlena pro simulace, jež mají za cíl věrohodně modelovat fenomény na úrovni úkolů, episod a krátkých příběhů. *Úkolem* či

[3] Tento příspěvek vznikl rozšířením článku [4].

episodou se rozumí určité dění, které se odehrává v určité oblasti, zahrnuje jednoho či více virtuálních lidí, trvá řádově několik minut a většinou tvoří logický celek. *Krátkým příběhem* se rozumí několik úkolů či episod v rozsahu řádově desítek minut až hodin odehrávajících se v jedné či více oblastech. Příkladem úkolu je zalévání zahrady, příkladem episody je hospodská rvačka. Příkladem krátkého příběhu je dění během otevírací doby hospody nebo během procházení úrovně z počítačové hry.



Obrázek 1. Ukázka IVE – restaurace. Číšník obsluhuje dělníky.

2. Úroveň abstrakce a diskretisace světa. Základní, atomární úroveň simulace je úroveň *performativních akcí* (dále akcí). Akcí člověka v přirozeném reálném světě se rozumí proces, který člověk vykonává, který trvá řádově několik vteřin a jenž většinou interpretujeme jako uzavřený celek. Slovo „performativní“ znamená, že na akcích se zajímáme o vnější, behaviorální stránku aktorů; nikoli tedy o jejich vnitřní – fysiologické nebo mentální – pochody. Během provádění akce může člověk pracovat s objekty nebo se může přemisťovat z místa na místo. *Objektem* se rozumí předmět velikosti řádově centimetrů až desítek centimetrů. *Místem* se rozumí prostor o velikosti řádově metrů čtverečních. Místa jsou representována diskrétně. Příkladem akce je zvednutí předmětu, krok apod. Objektem je například ořezávátko nebo lednice. Místem je třeba prostor před lednicí. Veliké předměty jako například strom se chápou jako topologické zvláštnosti míst. Místnosti a stavby se chápou jako oblasti.

3. Dynamický aspekt světa. Rozumí se, že virtuální svět je dynamický a z pohledu jednotlivých bytostí nepředvídatelný (vlastnosti světů přejímáme podle [15]).

4. Učení a emoce. Je abstrahováno od učení; výjimku tvoří pouze pamatování si a zapomínání pozic objektů. Architektura IVE rovněž nepracuje s umělými emocemi.

Zmíněná úroveň abstrakce je běžná v simulacích týkajících se virtuálních bytostí. Klíčové je, že sekvence akcí virtuální bytosti by měly na zvolené úrovni abstrakce věrně modelovat sekvence akcí skutečné bytosti. Můžeme tomu říkat „episodická plausibilita“. Akce by například měly být prováděny v odpovídajícím kontextu (lidé pijí pivo v hospodě, ale ne na úřadě; když pivo dojde, kývneme na číšníka) a měly by být smysluplně řetězeny (když se napijeme piva, obvykle pak odložíme sklenici; neodcházíme, dokud není dopito). Naproti tomu pokud je v modelu zavedena vlastnost typu »hlad« nebo »chuť-pít-pivo«, nemusí tato ničemu skutečnému odpovídat. Tyto vlastnosti hrají v modelu roli toliko podpůrných proměnných, které mohou být využity k věrohodnému generování akcí.

3 Representace ISMA

Jedním ze základních problémů při modelování virtuálních bytostí je representace znalostí o tom, co a jak lze ve virtuálním světě provádět. Vzhledem k úrovni abstrakce a vytyčeným cílům je prakticky jediným možným řešením symbolický přístup „shora–dolů“. Ovšem jak přesně má taková symbolická representace vypadat? Intuitivní přístup, kdy jsou popsány objektivní vlastnosti objektů jako je „hmotnost“, „tvar“, „absolutní směr pohybu“ atd. a vztahy mezi objekty jako „je součástí“, selhává. Taková representace totiž nevhodně předjímá způsob, jakým virtuální bytosti budou vnímat svůj svět – popis je příliš „nízkoúrovňový“ vzhledem ke zvolené abstrakci. Abychom vyřešili tento problém, navrhli jsme vlastní representaci. Lze ji označit za „relativně symbolickou“, neboť narozdíl od předchozího typu representace popisuje virtuální svět relativně vzhledem k bytostem, jež svět obývají.

Representaci označujeme ISMA: *intence – vhodnost – činnost – rada* („intention – suitability – materialization – advice“). Při jejím návrhu jsme se inspirovali Gibsonovou *teorií afordancí* („affordances“) [10]. Jedná se o percepční teorii, která bývá označována jako ekologická v tom smyslu, že popisuje vnímání organismu na úrovni jeho interakce s prostředím. V jistém smyslu abstrahuje kognitivní procesy, přesněji zahrnuje je do procesů percepčních, což představuje přesně tu úroveň abstrakce, kterou potřebujeme pro episodické simulace. Dále jsme vyšli z určitých prvků Bratmanovy teorie praktického rozhodování [1] (která je v komunitě zabývající se softwarovými agenty známá spíše ve své komputační variantě pod názvem *Belief Desire Intention*).

Snad proto, že jsme aplikaci pojali komplexně a umožňujeme modelování různých typů organismů a různých výřezů přirozeného reálného světa – narozdíl od většiny podobných aplikací – nastaly dvě skutečnosti:

- Mimovolně jsme opustili zažitou dichotomii pracující s rozhraním organismus–svět. O určitých prvcích virtuálního světa přestává mít smysl hovořit jako o prvcích uložených „ve světě“ nebo „v hlavách virtuálních lidí“. Tyto prvky lze chápat jako odkazující k určitým entitám skutečného světa, jež nejsou redukovatelné na entity náležející ani čistě do prostředí, ani čistě k aktorům (tedy jejich tělům nebo mentálním procesům).

- Ukázalo se, že přijatou representaci, jež původně precizovala nekomputační teorii afordancí, lze zpětně interpretovat jazykem soudobé ekologické psychologie, konkrétně následovníků Gibsona, kteří se zaobírají významem pojmu afordance ([17], [19]). Tím došlo k interpretaci čistě technického konstruktu – datových struktur representujících rozličné entity virtuálního světa – jazykem vědy, která si klade i filosofické otázky. Díky tomu můžeme ISMA representaci chápat až metafyzicky.

V tomto článku se budeme zabývat bodem prvním. IVE nám poskytne informace o tom, jestli je metodologicky výhodnější (z technického hlediska, na dané úrovni abstrakce) chápat rozhraní organismus–svět jako ostré, nebo spíše chápat organismy–svět jako spřažený systém. Dále nás dovede k pojmu „sdílená intence“ a nabídne přirozené vysvětlení významu tohoto pojmu. Bodem druhým se částečně zabýváme v článku [5].

3.1 Prvky nehierarchické ISMA

Representace ISMA má hierarchickou povahu, která umožňuje škálovatelnost simulace pomocí level-of-detail techniky. Pro snazší vhled do problému začneme popisem entit ISMA nehierarchickým způsobem a poté přejdeme na hierarchický popis. Podotýkáme, že na základě předkládané representace pracuje IVE.

První věc, kterou musí representace znázorňovat, je vazba mezi akcemi a objekty, které mohou v akcích vystupovat. Tato vazba je representována prostřednictvím tzv. *poukazů* a *slotů*. Representaci akcí se říká *činnost*.

Definice 1. *Objekt* („object“).

Objekt je abstraktní entita representující určitý objekt z přirozeného reálného světa. Každý objekt může mít *vlastnosti* odpovídající modelovaným vlastnostem vzorového objektu – většinou půjde o objektivní vlastnosti jako je hmotnost ap. Dále má objekt *poukazy*. Poukazem se rozumí reference na slot činnost.

Definice 2. *Činnost* („materialisation“).

Činnost je abstraktní entita virtuálního světa representující akci, kterou lze ve virtuálním světě provést; to jest podmínky, za kterých lze akci provést, a její efekt. Každá činnost může být prováděna s různými typy objektů. Pro každý typ objektu činnost disponuje slotem. Na sloty poukazují konkrétní objekty.

V tuto chvíli si představujme, že všechny činnosti trvají stejnou jednotku simulárního času. Z implementačního hlediska lze činnost vidět jako proceduru z vyššího programovacího jazyka, jejíž kód mění virtuální svět ve smyslu modelované akce. Sloty jsou jako parametry, za které je možné dosadit jednotlivé proměnné – objekty. O objektech lze pak hovořit jako o entitách, které „nabízí“ činnosti, na něž poukazují.

Činnost může provést buď virtuální člověk, nebo může být spuštěna sama. V obou případech ale musí být *instanciovaná*, to jest musí být jasné, s jakými objekty má být provedena. Činnosti chápeme jako entity s podobným statutem existence jako mají objekty – jsou součástí virtuálních světů.

Další vazbu, kterou potřebujeme zachytit, je vazba mezi akcemi, virtuálními lidmi a jejich záměry a místem, kde virtuální lidé mohou akce provádět.

Definice 3. *Aktor* („actor“).

Aktorem se rozumí objekt, který reprezentuje tělo virtuálního člověka. Aktor má určité *vlastnosti* a dále *poukazuje* na sloty právě těch činností, které virtuální člověk může vykonávat.

Definice 4. Génius („genius“).

Génius je řídící algoritmus virtuálního člověka.

V tzv. *základním případě* lze virtuálního člověka nahlížet jako dvojici <aktor, génius>. Později uvidíme rozšíření základního případu. Podotýkáme nicméně, že zatímco aktor modeluje tělo člověka, génius nemodeluje mysl. Génie si pro tuto chvíli představujme toliko jako funkce sloužící k výběru akcí tak, aby simulace byla věrohodná na úrovni episod a krátkých příběhů. A zatímco tělo virtuálního člověka na této úrovni alespoň zjednodušeně modelovat potřebujeme (uživatel simulace ho uvidí), mysl nikoli (uživatel ji neuvidí).

Dále potřebujeme representovat prostor virtuálního světa.

Definice 5. *Místo* („way-place“) *a topologie* („topology“).

Topologie je abstraktní representace prostoru virtuálního světa. Jde o multigraf, jehož uzly jsou místa a orientované hrany jsou přechody mezi místy.

Místo je speciální typ objektu; diskrétní, dále nedělitelná abstraktní jednotka representující určité místo přirozeného reálného světa. Místo *sousedí* s okolními místy. Místo může mít vlastnosti odpovídající modelovaným vlastnostem (například povrch).

Každé místo nabízí aktorovi stání a možnost na místo přejít z jiných míst. Speciální místa mohou nabízet různý typ pohybu, případně různý typ pohybu pro různé aktory. Předpokládáme, že každý objekt se nachází na nějakém místě nebo v/na nějakém jiném objektu.

Místa připomínají way-pointy z akčních počítačových her. Narozdíl od way-pointů, které jsou většinou používané pouze pro hledání cesty, místa representují „fyzikální“ prostor virtuálního světa. Není předepsáno, jestli topologie je 2D šachovnicová, pseudo-3D apod. – to závisí na konkrétním světě. Různí aktoři mohou dokonce vidět prostor různě – to jest mohou mít různou topologii.

Většina činností má jeden slot vyhrazený pro aktory a jeden pro místo. Takové činnosti mohou provádět aktoři, kteří na činnost poukazují, na místech, která na činnost poukazují. Činnosti lze tedy chápat jako lokalizované entity; jejich poloha je dána množinou míst, která na činnost poukazují. Mohou existovat i činnosti s více sloty pro aktory, s více sloty pro místa a bez slotů pro aktory.

V každý okamžik může být virtuální člověk obklopen různými činnostmi a různými objekty, navíc může mít různé cíle. Které činnosti a jak instanciované by měl jeho génius spustit? K tomu potřebujeme zavést do architektury *intence* a *vhodnosti*.

Cíle, přání a intence („goals“, „desires“, intentions“) chápeme ve smyslu Bratmana [1]. Cíle jsou zde *budoucí stavy věcí* („state of affairs“). Člověk může mít *přání* dosáhnout určitého cíle (stavu světa). Člověk ovšem může mít různá přání, z nichž ne všechny nutně povedou k jednání. Může mít přání dát si koláček, ale jelikož je tlustý a drží dietu, nedá si

ho. Naproti tomu *záměr dosáhnout cíle* (neboli *intence dosáhnout cíle*) označuje stav, v němž se člověk k vykonání nějakého přání pevně rozhodl. Přestože je tlustý, může se rozhodnout, že si odpoledne ke svačině jeden koláček dá. Jakmile jednou záměr *přijal* („commit"), nebude už za normálních okolností znovu vážit pro a proti a až přijde čas, koláček sní.

Bratman ve své knize rozpracovává metodologii hodnocení racionality lidského chování v závislosti na tom, jak lidé přijímají záměry a jak se jich vzdávají. Například člověk není (podle metodologie) racionální tehdy, když přijal intenci držet dietu a zároveň jíst koláčky. Bratman rozpracovává dále vlastnosti intencí, například že přetrvávají a že ovlivňují budoucí rozhodování. Tyto aspekty jeho práce pro nás nejsou podstatné. Klíčové pro nás je, že Bratman identifikoval (mezi jinými) dva základní typy intencí: intence *dosáhnout cíle* a intence *dosáhnout cíle určitým způsobem*. V ISMA budeme pojem intence chápat v prvním smyslu – jako záměry dosáhnout určitých cílů. Činnosti pak tedy představují způsoby, jak cílů dosáhnout. Ovšem jelikož způsobů může být více (to jest více činností může vést ke splnění téhož cíle), je potřeba se pro nějakou konkrétní činnost rozhodnout. Tedy v Bratmanově terminologii přijmout intenci dosáhnout cíle určitým způsobem. Toto rozhodnutí bude činěno na základě *vhodností*. Vhodnosti jsou v ISMA funkce určující, která činnost pro dosažení dané intence má být provedena. Intence mají hierarchickou povahu. Tím se rozumí to, že k dosažení určité intence může být nutné přijmout později další intence. Když chceme koláček, musíme si ho napřed koupit. Zatím zavedeme intence nehierarchické.

S přáními ve smyslu Bratmana zde pro zjednodušení pracovat nebudeme.[4]

Definice 6. *Intence* („intention").

Adoptovaná intence neboli *záměr* je abstraktní entita representující stav připravenosti virtuálního člověka dosáhnout nějakého cíle v současnosti nebo v budoucnu. Každá intence *poukazuje* na jednu či více činností, které vedou k jejímu dosažení. Adoptovaná intence může být:

- *aktivní*; odpovídá připravenosti konat akce nyní, přičemž v danou chvíli přímo *je* vykonávána nějaká činnost z těch, na něž intence poukazuje,
- *preaktivní*; určuje připravenost konat akce nyní, přičemž ale žádná činnost z těch, na které intence poukazuje, vykonávána v tento okamžik není,
- *neaktivní*; určuje *připravenost* konat akce v budoucnu.

Potenciální intence je abstraktní entita, kterou génius může přijmout jako adoptovanou intenci, ale zatím tak neučinil. Potenciální intence tedy representuje potenciálně možný stav připravenosti virtuálního člověka konat určité akce v budoucnu.

Definice 7. *Vhodnost* („suitability").

Vhodnost je funkce náležející určité činnosti, která na základě vlastností objektů poukazujících na sloty činností (včetně aktorů) a případně dalšího kontextu určí, jak vhodné je v danou chvíli pro daného virtuálního člověka s danou intencí provést činnost s danými objekty (výstup funkce lze chápat např. jako <0, 1>).

[4] Komputační varianta této teorie, tedy architektura BDI, sice s přáními pracuje („desires"), viz např. [21], avšak v posunutém významu. Přání se zde chápou víceméně jako potenciální intence, nikoli specifické stavy mysli resp. postoje zaujímané vůči cílům.

Nyní je možné říci, co je to *smysl objektu pro aktora* – tedy sémantický význam objektu. Význam objektu pro daného aktora je tvořen dvěma složkami. První složku bychom mohli nazvat *disposiční*. Tvoří ji právě ty poukazy objektu, které vedou na sloty právě těch činností, jež může aktor potenciálně provádět (tedy poukazuje na ně). Jde tedy o jakýsi obecný, „objektivní“ význam objektu nezávislý na kontextu. Druhý aspekt významu, mohli bychom ho nazvat aktuální složkou významu, je představován zhodnocením vhodnosti použít daný objekt určitým způsobem (to je na daném slotu) při určité činnosti v aktuální situaci.

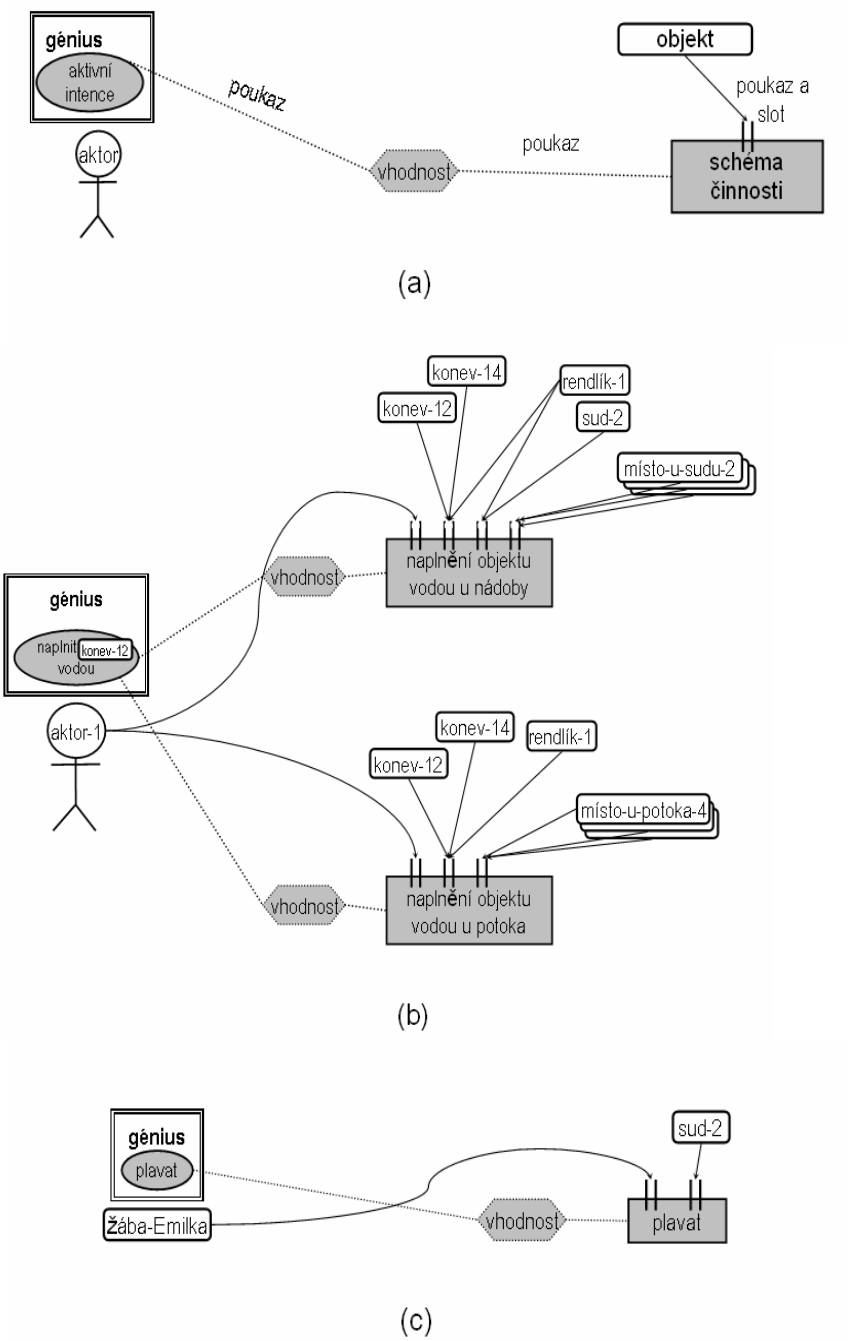
Uvědomme si, že oba aspekty významu mají relační povahu; různé objekty mají různý význam pro různé aktory. Ukážeme to na následujícím příkladu: Uvažme jablko. Disposiční význam jablka pro člověka může být, že se s ním dá házet a že se dá jíst, přičemž v obou případech potřebujeme ještě ruku. Aktuální význam jablka pro určitého konkrétního člověka v daném kontextu může být, že je lepší k jídlu než fazolky na smetaně, ale horší, než guláš (z definice se předpokládá, že v dané situaci jsou fazolky, guláš i jablko k disposici). Disposiční význam jablka pro srnu může být, že se dá sežrat.

Sedm prvků – objekty, aktoři, místa, činnost, vhodnosti, intence a géniové – tvoří společně s poukazy mezi entitami stavební kameny ISMA representace. Několik situací ilustruje obrázek 2.

V tuto chvíli tedy máme danou základní strukturu pro vytváření representací virtuálních světů. Jakým způsobem se ale může virtuální člověk rozhodovat na základě takové representace? Odpověď je skryta ve způsobu, jakým génius virtuální svět vnímá. Existuje několik způsobů vnímání, které representace přirozeně umožňuje. Zde si vezměme dva základní typy:

- *Aktivní vnímání objektů a míst.* Při tomto typu vnímání se génius „rozhlédne“ po svém okolí (pro jednoduchost předpokládejme, že okolí je nějak věrohodně určeno – v IVE se za tím účelem pracuje se *senzory*) a skrze poukazy svých aktivních intencí nalezne činnosti vedoucí k uspokojení jeho cílů. Poté se pustí proti směru poukazů objektů na dané činnosti – a nalezne příslušné objekty a místa. Při tomto typu vnímání génius ignoruje vše, co nesouvisí s tím, co právě chce dělat.
- *Aktivní vnímání možností jednat.* Při tomto typu vnímání se nejprve génius „rozhlédne“ po svém okolí a skrze poukazy svých aktivních intencí nalezne činnosti vedoucí k uspokojení jeho cílů. Poté pomocí předchozího typu vnímání nalezne objekty a místa, kterými by mohl instanciovat činnosti. Nedeterministicky instanciuje činnosti a použije jejich vhodnosti, aby zjistil, jak je výhodné s ohledem na daný kontext začít tu kterou instanciovanou činnost vykonávat.

Analogicky by bylo možné hovořit o pasivním vnímání, při kterém by si génius mohl samovolně něčeho všimnout a které by tak umožnilo modelovat reflexivní chování. Bylo by možné hovořit i o vnímání ovlivněném preaktivními či neaktivními intencemi. Lze také model rozšířit o paměť – génius by se pak nerozhlížel pouze po okolí, ale i hledal informace v paměti [13]. Zde však budeme pracovat pouze s aktivním vnímáním.



Obrázek 2. Část (a) schématicky zachycuje jednotku ISMA. V části (b) vidíme dvě jednotky ISMA, které mají společnou intenci „naplnit objekt vodou“. Jelikož virtuální člověk má tuto intenci aktivní, dá se předpokládat, že v ruce drží konev (i když to obrázek neznázorňuje). Vidíme, že v danou chvíli jsou dostupné tři možnosti, jak konev naplnit – v sudu, z rendlíku a u potoka. Vidíme, že v danou chvíli lze teoreticky naplnit ještě další konev a rendlík. Vhodnosti určí, jaká kombinace je nejvýhodnější – pravděpodobně půjde o naplnění konve, kterou aktor drží, a to u sudu nebo u potoka. ISMA v části (c) znázorňuje, že žába Emilka může v témže sudu plavat (což virtuální člověk nemůže). Sud pro žábu není objektem, ale místem. Žába tedy vidí jinou topologii než virtuální člověk.

Pokud tedy předpokládáme, že génius má nějaké adoptované aktivní intence, je pro něj jednoduché pomocí ISMA representace rozhodnout, jaké akce má provádět. Jednoduchost je zde míněna z technického hlediska – poukazy umožňují efektivní vyhledání žádaných činností a objektů. Prostředí totiž tyto díky representaci přímo nabízí. Representace zároveň řeší druhý z cílů projektu, totiž rozšiřitelnost. Pokud přidáme objekt s poukazy, automaticky bude objekt nalezen během vnímání. Analogicky to platí pro přidání nové činnosti, pokud se váže k existující intenci.

Vnímání v IVE funguje tak, jak bylo popsáno. Přestože je ISMA zamýšlena jako technický konstrukt pro episodické simulace, podle ekologické psychologie ji navíc lze považovat za psychologicky plausibilní (připomínáme, že jsme abstrahovali nad úrovní kognitivního procesu). Rozbor tohoto tématu ale přesahuje rámec článku.

3.2 Prvky hierarchické ISMA

Nejprve zavedeme do architektury *oblasti*, pomocí nichž půjde representovat hierarchicky topologii. Poté přejdeme k hierarchické representaci činností a intencí.

Definice 8. *Oblast* („area“) a *topologie* hierarchicky.

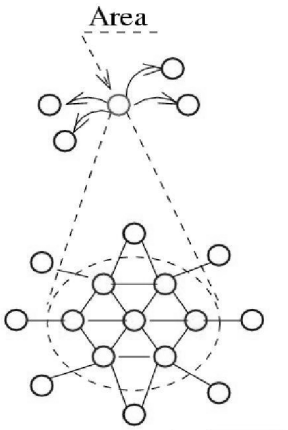
Oblast je entita representující určitou lokaci přirozeného reálného světa. Obsahuje buď podoblasti, nebo místa. Oblast sousedí s okolními oblastmi. Může mít určité specifické vlastnosti.

Topologie je hierarchický multigraf. Každá hladina multigrafu je sama o sobě multigraf a představuje určitou úroveň abstrakce prostoru. Nejnižší úroveň abstrakce je topologie míst podle def. 5. Nejvyšší úroveň představuje abstraktní bod representující celý virtuální prostor. Uzly ostatních multigrafů jsou oblasti a orientované hrany v rámci jedné úrovně jsou *přechody* mezi oblastmi. Z každého uzlu z jakékoli úrovně vyjma nejnižší vedou neorientované hrany též k určitým uzlům z nejbližší nižší úrovně. O těchto uzlech říkáme, že detailněji representují prostor oblasti a že jsou v ní *obsaženy*. Samy o sobě jsou podoblastí nebo místem. Každý uzel může mít takto právě jednoho otce. Každá oblast obsahuje alespoň jednu podoblast či místo.

Speciální místa a lokace mohou nabízet různý typ pohybu, případně různý typ pohybu pro různé aktory.

Každý objekt se nachází na nějakém místě nebo v/na nějakém jiném objektu a zároveň se nachází ve všech nadřazených lokacích.

Tento způsob representace virtuálního prostoru je poměrně přirozený, používá se například v počítačových hrách. Situaci ilustruje obrázek 3. Z technického hlediska umožňuje tato struktura například hledat cestu pomocí hierarchického A*. Všimněme si, že hranice prostoru jsou ostré. Tato representace vyhovuje spíše uzavřeným prostorám – místnostem, patrům, budovám. Méně se už hodí pro otevřené prostory, například pole.



Obrázek 3. Hierarchie oblastí. Oblast se rozpadá na podoblasti nebo místa. Převzato z [12].

Definice 9. *Činnost* hierarchicky.

Činnost je abstraktní entita virtuálního světa representující úkol nebo atomickou akci, kterou lze ve virtuálním světě provést.

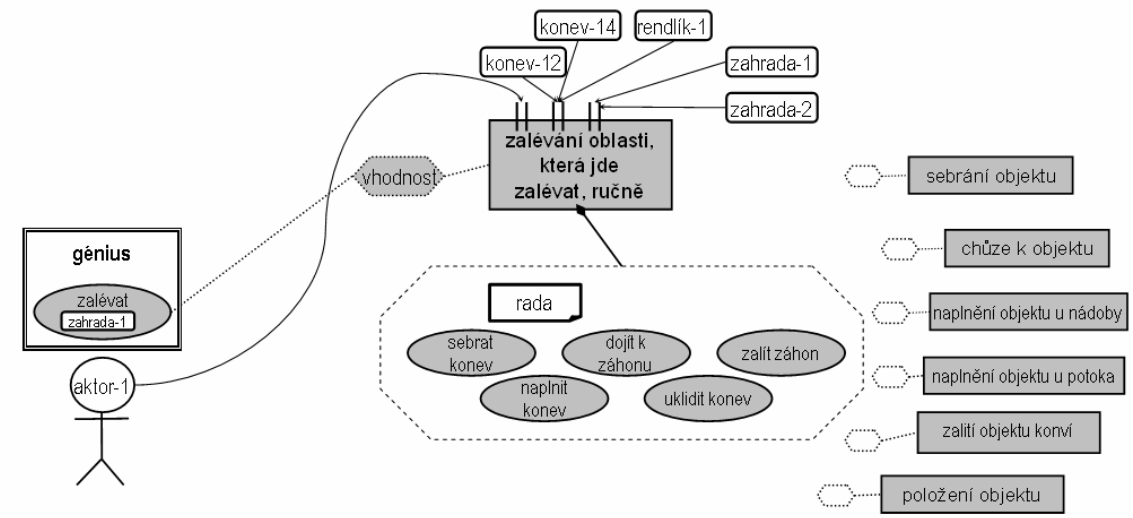
Pokud činnost representuje úkol, jedná se o trojici <*S*, *rozklad*, *suit*>, kde *S* je množina slotů, na něž mohou poukazovat konkrétní objekty, místa a oblasti, *suit* je vhodnost dle def. 7 a *rozklad* („breakdown“) je dvojice <*rada*, *Subintence*>. *Subintence* je množina potenciálních intencí, které by měl génius adoptovat, pokud chce vykonat dané činnosti. *Rada* je funkce, která kdykoli na základě aktuálního kontextu poskytuje géniovi jednoznačné informace o tom, kterou ze Subintencí v danou chvíli zvolit jako aktivní, které jako preaktivní a které jako neaktivní, případně o tom, jestli už úkol není splněn a jak dobře, případně jestli se situace nezměnila tak, že už úkol nelze splnit.

Pokud činnost representuje akci, jedná se o trojici <*S*, *suit, efekt*>, kde *S* je množina slotů, *suit* je vhodnost a *efekt* je popis transformace světa – to je v souladu s původní definicí (def. 2). Takové činnosti budeme označovat jako *atomické* („atomic“).

Význam rozkladu je následující: Jakmile se génius aktivně rozhlédne, aby spatřil možnosti jednat ve prospěch určitého cíle, dozví se o nejvhodnější činnosti. Ta mu prostřednictvím rozkladu nabídne intence, které by měl adoptovat, aby cíle dosáhl, a dále mu poskytne radu, v jakém pořadí má intence aktivovat v závislosti na aktuálním kontextu. Rozklad je tedy návodem od prostředí pro génia o tom, jaké intence adoptovat.

Jelikož nabízené intence poukazují na podřazené činnosti, jež mohou nabízet další subintence, lze chápat činnosti a intence jako hierarchicky organizované. Struktura nemusí být čistě stromová, protože pokud máme dva cíle, některé činnosti, které vedou k jejich dosažení, mohou být stejné. Na nejvyšší hladině se nachází hlavní intence – tedy záměry dosáhnout hlavních cílů. Ty nejsou poskytované prostředím – génius je má od počátku a aktivuje je podle svých potřeb.

Jako příklad uvažme opět zalévání: podintencemi zalévání zahrady konví mohou být sehnání konve, naplnění konve vodou, dojití ke konkrétnímu záhonu, zalití toho záhonu a uklizení konve. Prostřední tři mohou být aktivní opakovaně. Viz obrázek 4.

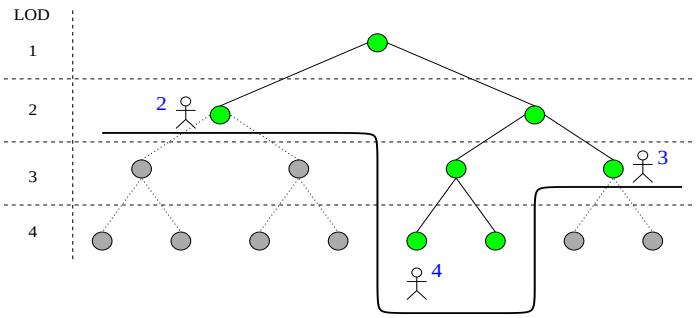


Obrázek 4. Hierarchická ISMA. Intence zalévat konkrétní zahradu poukazuje na činnost zalévání konví (druhou činností by mohlo být třeba zalévání hadicí). Zalévání pak poukazuje na rozklad, který obsahuje potenciální intence.

I nyní si budeme představovat, že všechny atomické činnosti trvají stejnou jednotku simulárního času. Ostatní činnosti mohou trvat podle aktuální situace. Pokud předchozí den pršelo, zalévání půjde rychle. Pokud nepršelo a nejbližší voda je ve dva kilometry vzdáleném potoku, půjde pomalu.

Opět podotýkáme, že se nejedná o „prázdnou“ definici. Takto to v IVE funguje. Příklad se zaléváním je sice převzat z jiného projektu – z projektu Enti [2] – je ale, až na „dva kilometry vzdálený potok“, rovněž implementován [7]. Poněkud abstraktní rady a vhodnosti byly implementovány pomocí reaktivních fuzzy rozhodovacích pravidel.

Z technického hlediska je nyní podstatné, že máme dvě hierarchie abstrakcí – prostorovou a procedurální – jejichž úrovně spolu určitým způsobem korespondují. To umožnilo použít v IVE techniku level-of-detail pro řízení virtuálních lidí. Virtuální lidé, kteří se nachází na „nepodstatných místech“, nebudou provádět činnosti na nejnižší úrovni abstrakce, nýbrž na vyšších hladinách. Viz obr. 5.



Obrázek 5. LOD v IVE. Některé postavy jsou simulovány detailně, jiné na vyšších úrovních abstrakce.

Representace tak přirozeně pomohla vyřešit první z cílů projektu. Technika level-of-detail je popsána v [14].

4 ISMA a rozhraní organismus–svět

Až doposud jsme se bavili o tom, „jak to v IVE je“. Nyní jsme dospěli k bodu, kdy můžeme diskutovat representaci ISMA v širším kontextu. Zaprvé se pokusíme pomocí ISMA definovat pojem virtuální svět a budeme argumentovat proti zažité dichotomii pracující s rozhraním organismus–svět. Zadruhé rozebereme pojem sdílené intence, jenž se přirozeně objeví při hledání rozhraní.

V této sekci půjde především o definici virtuálního světa a hledání hranice mezi organismem a prostředím. Předjímáme, že klíčová pro toto hledání bude odpověď na otázku: Co modeluje génius? Tuto otázku v rámci sekce zodpovíme.

Hierarchická ISMA nám dává o virtuálním světě následující obraz:

Definice 10. *Virtuální svět.*

Virtuální svět je dvojice <**Γ**, **Π**>, kde **Γ** je sféra géniů a **Π** fyzická část virtuálního světa.

Γ je čtveřice < I_{ALL}, G_{ALL}, M_{ALL}, B_{ALL} >. I_{ALL} je množina všech potenciálně adoptovatelných intencí včetně poukazů na příslušné činnosti. G_{ALL} je množina géniů včetně jejich vlastností, paměti a adoptovaných intencí a vazby na ovládaného aktora. B_{ALL} je množina všech

rozkladů činností. Každý rozklad obsahuje radu a poukazuje na několik potenciálně adoptovatelných intencí. M_{ALL} je množina činností s jejich vhodnostmi a poukazem na jejich rozklad.

Π je dvojice $<O_{ALL}, WA_{ALL}>$, kde O_{ALL} je množina objektů včetně jejich vlastností a poukazů a WA_{ALL} topologie.

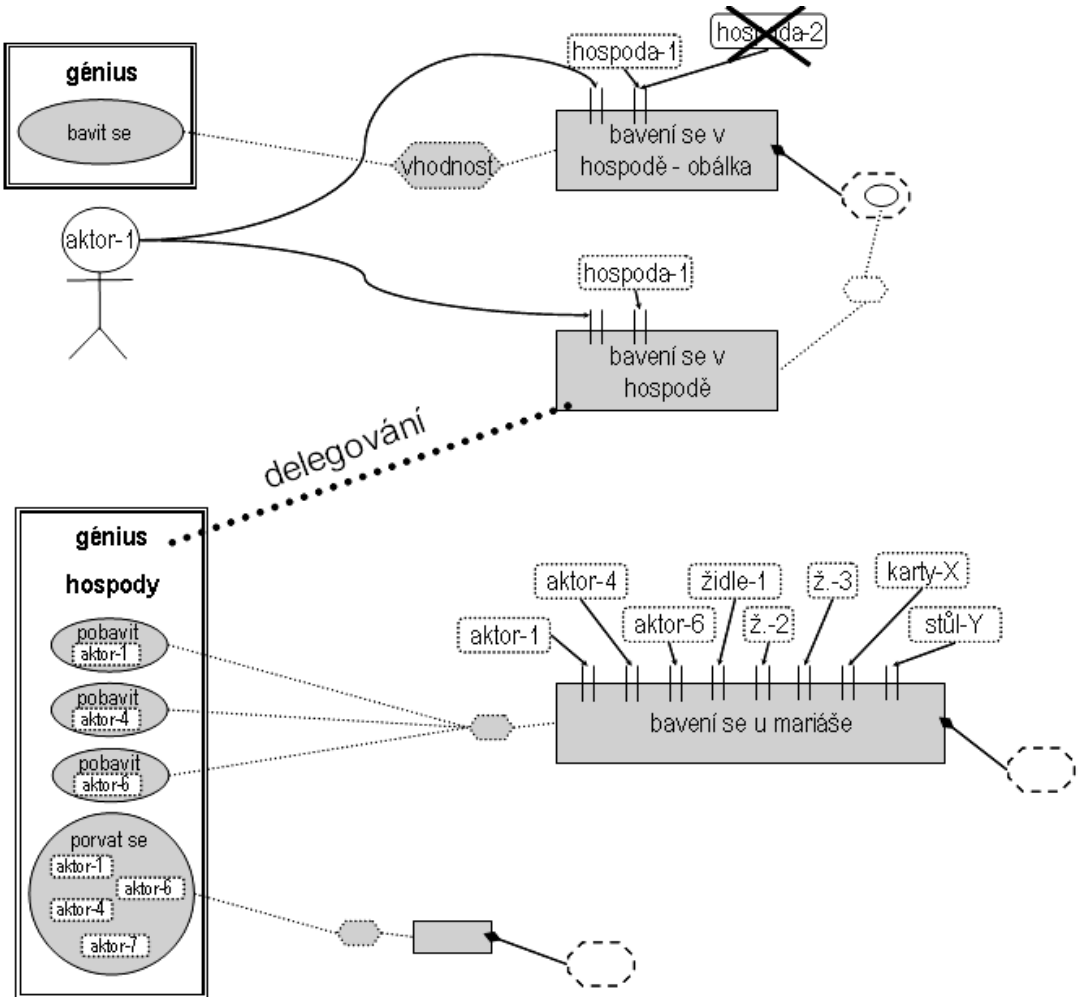
Virtuální svět podle této definice tedy zahrnuje jak virtuální lidi, tak jejich prostředí. Ptáme-li se po hranici organismus–svět, ptáme se po rozhraní mezi virtuálními lidmi a prostředím. Existuje takové rozhraní? Intuitivně lze říci, že ho tvoří vhodnosti, přičemž činnosti a objekty patří do světa a aktoři a géniové utváří „organismy“. Domníváme se, že tento pohled je v zásadě správný, nicméně nepodstatný. Celá otázka po hranici organismus–svět je totiž nevhodně položena.

Z implementačního hlediska se jistě lze ptát po rozhraní mezi komponentami systému. Pak je pochopitelně možné hovořit (technickým jazykem) o rozhraních mezi aktory a génii, mezi aktory a objekty, mezi génii a činnostmi a podobně. Ovšem to, co nás nyní zajímá, nejsou hranice v technickém slova smyslu. Zajímá nás, jestli representace v IVE změní naši intuitivní představu o hranici mezi organismem a jeho prostředím v přirozeném reálném světě. Protože to by nás mohlo vést k tomu, že napříště budeme modelovat přirozený svět odlišným způsobem. Budeme argumentovat v tom smyslu, že chápeme-li přirozený svět na úrovni abstrakce IVE, to jest popisujeme-li ho na úrovni akcí a objektů, dochází k „rozmazání“ hranice mezi prostředím a organismem. Vyvstane ovšem jiná hranice; hranice mezi statickou a dynamickou složkou světa.

Tento závěr přirozeně vyplyne z otázky: Co vlastně modeluje génius? Říkali jsme, že nikoli kognitivní procesy. Proč ne? Vždyť přeci hovoříme o tom, že adoptuje intence, rozhlíží se a má paměť.

Ukážeme několik příkladů, k čemu byli géniové v IVE použiti:

1. Génius v základním případě řídí chování jednoho aktora přesně tak, jak bylo popsáno výše.
2. Génius řídí vozík v dole stejným způsobem, jako aktora. Mezi vozíkem a aktorem není z technického hlediska příliš veliký rozdíl.
3. Génius řídí aktora, který něco drží, například kámen (nebo obecně nástroj): génius tedy řídí systém aktor–instrument.
4. Speciální géniové pomáhají základnímu géniovi řídit aktory v situaci, kdy je aktorů více pohromadě. Příkladem je hospoda. Základní génius, než vejde do hospody, „požádá“ speciálního génia hospody, aby mu pomohl se řízením aktora. Génius hospody „vidí celou hospodu“, může tedy řídit aktory centrálním způsobem. Pokud ale tento génius určí, že si má dojít aktor pro pivo k barovému pultu, předá ho u pultu dalšímu géniovi – géniovi fronty. V tuto chvíli tak řídí jednoho aktora tři géniové najednou. Původní génius přitom například stále kontroluje, jestli aktor nepotřebuje na toaletu.
5. Přestože jsme toho v IVE nevyužili, mohou speciální géniové mít jakési „sdílené“ intence; například génius hospody může mít intenci vyvolat rvačku. Obrázek 6 to zachycuje.



Obrázek 6. Génius hospody má sdílenou preaktivní intenci způsobit rvačku. Aktivní intence se týkají hraní mariáše. Ačkoli tyto aktivní intence spravuje génius hospody, dají se chápat i jako individuální intence aktorů č. 1, 4 a 6 (narozdíl od intence „porvat se“).

Má v případě 2, 4 a 0 smysl klást si otázku, čí kognitivní proces je modelován a kde je hranice mezi virtuálním člověkem a prostředím? Cenné jsou hlavně případy 4 a 0. Je v případě 3 držený nástroj součástí virtuálního člověka nebo prostředí? Tyto otázky podle nás nejsou klíčové. Klíčová je otázka, co dělají géniové. Pokud identifikujeme to, co mají géniové společné, zjistíme, že *géniové řídí systémy – jsou modely jejich dynamiky*.

Z technického hlediska se nám tedy ukázalo, že je výhodné (na dané úrovni abstrakce) pojímat přirozený svět jako fyzikální entity a systémy (systémy v intuitivním slova smyslu). Tyto systémy jsou dynamické. To znamená zahrnují fyzikální entity, mohou pojmout novou entitu či z nich může být určitá entita oddělena. Dále určují, kdy budou fyzikální entity změněny prostřednictvím činností. Systémy mohou průběžně vznikat a zanikat (uvažme frontu) a mohou se mezi sebou navzájem překrývat.

Hranice, jež tedy vede virtuálním světem, je hranice mezi jeho statickou a dynamickou složkou. Je to hranice mezi **Γ–Π**. Z technického hlediska se pak jeví jako výhodné vytvářet génie pro všechno, co se nám jeví jako dynamický systém. Pochopitelně

hranice mezi tím, co je systém a co už není, i mezi jednotlivými systémy, závisí na interpretaci návrháře světa.

Diskutovali jsme sice o technických konstruktech (génius je stále pouze entita representace), nicméně je snad možné učinit následující úvahu: Jelikož IVE je modelem určité části přirozeného reálného světa a jelikož je to právě representace ISMA, která zachycuje tuto část, a jelikož v ISMA je výhodné chápat hranici mezi její dynamickou a statickou složkou – není snad výhodné chápat takto i onu modelovanou část přirozeného reálného světa na dané úrovni abstrakce? Vždyť co jiného je nějaká terminologie, v níž chápeme nějaké jevy, než representace těchto jevů?

5 Sdílené intence

V příkladu 0 jsme použili pojem intence v nezvyklém významu – totiž ve významu *sdílené intence* („joint intention"). K tomuto bodu se nyní vrátíme. Je možné hovořit o tom, že hospoda má intenci vyvolat rvačku? Na to je možné odpovědět dvěma způsoby. Za prvé, pokud vidíme hospodu jako agenta v technickém smyslu (tedy například dle Wooldridge [21]), pak samozřejmě ano. Jestli může mít intence BDI agent filtrující spamy, proč by je nemohl mít agent řídící dění v hospodě? Jenomže to nám o povaze sdílené intence nic neříká. Více se dozvíme při pokusu odpovědět na otázku: *Co sdílená intence modeluje?*

Pokud rozšíříme definici 6, můžeme říci, že *sdílená intence modeluje stav připravenosti určitého systému přejít do nějakého stavu v současnosti nebo v budoucnu*. To intuitivně říká například to, že pokud má génius hospody intenci vyvolat rvačku, znamená to, že „se schyluje ke rvačce". Intuitivně tedy intence odráží to, že aktuální stav, v němž se hospoda nachází, pravděpodobně v brzké době přejde do dalšího stavu – „rvačka". Při tom samozřejmě opouštíme Bratmanův myšlenkový aparát na posuzovaní racionality rozhodování: nemá smysl hovořit o tom, že hospoda je nebo není racionální v závislosti na tom, jestli rvačka skutečně nastala nebo ne. Mělo by samozřejmě smysl precizovat pojem „připravenost". Do toho se zde ale nebudeme pouštět – intuitivně si představujme, že *připravenost systému v daném okamžiku vůči něčemu* znamená, že pokud by byl systém uzavřený, velmi pravděpodobně by po určité době skončil ve stavu, vůči němuž je připraven.

Tato úvaha má praktickou motivaci, související s následující otázku: Kdy je výhodnější representovat připravenost systému sdílenou intencí a kdy ji redukovat na intence komponent systému (zde géniů řídících jednotlivé aktory, kteří by se měli poprat)? Konkrétně: mají mít jednotliví géniové intenci se poprat nebo je výhodnější zavést intenci sdílenou?

Domníváme se, že odpověď souvisí s tím, jestli lze existenci sdílené intence odvodit („infer") z jiných entit modelu ISMA. Tedy vztaženo na přirozený svět: jestli prvky, z nichž lze sdílená intence odvodit, jsou prvky alespoň ze základní úrovně abstrakce ISMA – tedy nikoli úrovně subsymbolické (v intuitivním významu toho slova). V tomto kontextu: je připravenost ke rvačce vidět (při abstrakci na úrovni objektů a atomických akcí) na nějakých konkrétních individuálních virtuálních lidech? To si musí návrhář rozmyslet v závislosti na tom, jakou rvačku chce modelovat. Pakliže má jít o rvačku vyvolanou jedním člověkem, pravděpodobně je připravenost redukovatelná na jeho soukromou intenci nebo vlastnosti. Možná je ten člověk opilý, možná přímo přichází se

záměrem se poprat. Pakliže ale má jít o rvačku vzniklou „emergentně“, může se stát, že připravenost sice redukovatelná je, ale na prvky „pod“ základní úrovní abstrakce. V tomto případě je třeba modelovat připravenost jako sdílenou. V prvním případě je možné se rozhodnout.

Ukážeme to ještě na názornějším příkladě: Představme si trojici, která hraje karty. Karty mají jasná pravidla, stavy hry jsou (i na naší úrovni abstrakce) diskrétní a dobře popsatelné. Mechanickou hru v karty (chápanou jako modelovaný jev), kde nikdo nepodvádí a nepere se, je možné namodelovat „zdola-nahoru“ – tedy každý génius může řídit svého aktora individuálně a rozhodovat se pouze podle svých lokálních informací (co drží za karty, co je vyloženo a co přišlo z balíčku).

A nyní si představme souboj Sandokana a tygra. Nejprve ti dva kolem sebe krouží, dívají se z očí do očí, sledují pohyby toho druhého – a pak na sebe v jednu chvíli, oba najednou, skočí. V systému Sandokan–tygr se „schyluje“ k tomu, že ti dva na sebe skočí, už chvíli předtím, než se tak stane. Tato připravenost ovšem nelze redukovat (na naší úrovni abstrakce) na žádnou vlastnost Sandokana ani tygra. Nonverbální pohyby, na jejichž základě oba poznají, že „teď je čas skočit“ jsou „subsymbolické“. Tuto situaci nelze namodelovat rozumně (na naší úrovni abstrakce) jinak než centrálně. Pokud by se podobná situace modelovala na naší úrovni abstrakce „zdola-nahoru“, vedlo by to pouze k tomu, že by se (technicky řečeno) dva agenti mezi sebou kradmo domluvili, kdy na sebe skočí. To je ovšem analogické tomu, když centrální génius rozhodne aktivovat příslušnou sdílenou intenci – jenom technicky zbytečně složité.

Podobně se lze ptát u rvačky v hospodě: vzniká daný typ rvačky, který chceme modelovat, spíše jako když se hrají mechanicky karty, nebo spíše jako v případě Sandokana a tygra?

Ukázali jsme tedy, že pojem sdílené intence dává smysl, i když nikoli v původním Bratmanově významu. Dále jsme poskytli metodologický návod – byť pouze orientační – k tomu, kdy je výhodné sdílené intence do virtuálního světa zanášet.

Na závěr ještě podotýkáme, že pojem sdílená intence se objevuje také v kontextu výzkumu kooperace mezi agenty v multiagentních systémech (např. [18]; [21], kap. 9). Z hlediska této oblasti je klíčové, jestli k určité akci v multiagentním systému dochází tím, že agenti a) spolupracují („coordinated cooperative action“) nebo b) nespolupracují („coordinated, but not cooperative action“) ([16], cit. z [21]). Někdy se v případě (a) o intenci dosáhnout týmového cíle (např. odnést ve čtyřech těžkou bednu) hovoří jako o intenci sdílené. Narozdíl od našeho chápání sdílené intence zde ale jde o intenci, kterou drží každý jednotlivý agent.

Pokud se vrátíme k našem příkladu se rvačkou: pokud vstoupí do hospody několik lidí se společným záměrem vyvolat rvačku, budeme to klasifikovat jako případ (a). Pokud bude člověk sám, budeme to charakterizovat jako klasickou intenci jednotlivce. Pokud rvačka vznikne na základě individuálního jednání několika opilců, kteří sice všichni chtějí rvačku vyvolat, ale nejsou nijak domluveni, dostáváme situaci (b). V tomto článku jsme navíc upozornili na čtvrtou možnost, která se od předchozích situací liší. Rvačka může také vzniknout spontánně, bez kooperace, ze situace, kdy je sdílená intence „rozprostraněna“ napříč celou restaurací a nedá se redukovat na intence jednotlivců, jinými slovy ze situace, kdy „rvačka visí ve vzduchu“.

6 Závěr

V článku jsme představili aplikaci IVE – nástroj na modelování velikých virtuálních světů – a representaci ISMA, se kterou nástroj pracuje. IVE jsme nepopisovali z implementačního hlediska, ani ho nevnímali jako to, čím fakticky je – platformou pro testování algoritmů umělé inteligence a prototypování rozsáhlých virtuálních světů typických pro počítačové hry. Místo toho jsme ho použili jako výchozí bod pro další diskuse o způsobu, jakým popisovat přirozený reálný svět na úrovni, na níž se s ním jako lidé běžně setkáváme – to jest na úrovni objektů a jednoduchých akcí.

Argumentovali jsme, že v okamžiku, kdy modelujeme veliký virtuální svět, tedy potažmo popisujeme větší část přirozeného reálného světa, dochází k přirozenému rozmazání hranice mezi organismem a prostředím – navzdory zachování zvolené úrovně abstrakce. Vyvstanou ovšem hranice nové: hranice mezi statickou a dynamickou „složkou“ světa, a pak dílčí hranice mezi systémy, jež dynamickou složku určují. Z technického hlediska to znamená, že je výhodnější modelovat systém jako jeden celek, nikoli oddělovat virtuální bytosti a prostředí jako dva typy „abstraktních strojů“.

Představili jsme dále nový koncept sdílených intencí, které se liší od sdílených intencí, s nimiž pracuje multiagentní výzkum. Sdílené intence jsme interpretovali jako „připravenosti“ systému přejít ze stávajícího stavu do jiného stavu. Od klasických sdílených intencí se liší tím, že nejsou redukovatelné na intence jednotlivců. Předložili jsme také orientační návod k tomu, kdy v simulaci se sdílenými intencemi pracovat a kdy pracovat s intencemi jednotlivců. Koncept sdílených intencí jenom podtrhuje nejasnost hranice organismus–prostředí.

Hlavní přínos příspěvku spatřujeme v tom, že poskytuje určité metodologické návody, jak výpočetně modelovat (na dané úrovni abstrakce) přirozený reálný svět; takové návody, které nejsou intuitivně nahlédnutelné, neboť překračují zažitou dichotomii organismus–prostředí.

IVE je k disposici na webovské stránce projektu [11].

Poděkování: IVE bylo naimplementováno v roce 2005 studenty MFF UK, Praha: O. Šerým, T. Pochem, P. Šafratou, J. Kubrem, J. Kulhánkem a Z. Šulcem, jímž tímto patří dík. Práce na tomto textu byla částečně podpořena grantovým projektem GA UK 351/2006/A-INF/MFF a grantem „Information Society“ pod číslem projektu: 1ET100300517.

Literatura

[1] Bratman, N.: *Intention, plans, and practical reason*. Cambridge, Mass: Harvard University Press 1987.

[2] Bojar, O., Brom, C., Hladík, M., Toman, V.: The Project ENTs: Towards Modeling Human-like Artificial Agents. In: *SOFSEM 2005 Communications.* Slovak Republic 2005, pp. 111–122.

[3] Brom, C., Lukavský, J., Šerý, O., Poch, T., Šafrata, P.: Affordances and level-of-detail AI for virtual humans. In: *Proceedings of Game Set and Match 2*, Delft 2006.

[4] Brom, C., Lukavský, J.: Proč je hranice mezi virtuální bytostí a jejím světem neostrá. V: *Sborník příspěvků z konference Kognice a umělý život*, Třešť. Opava: Slezská univerzita 2006, str. 97 – 106
[5] Brom, C., Lukavský, J.: Psychika Psychika virtuální bytosti. In: Sborník příspěvků z konference Kognice, Praha 2006, v tisku.
[6] Brom, C., Šisler, V., Hoffmann, M.: Virtuální lidé. *Umělá inteligence 5* (Mařík V., Štěpánková O., Lažanský J., eds), Praha, ČR: Academia 2007, v tisku.
[7] Brom, C.: Hierarchical Reactive Planning: Where is its limit? In: *Proceedings of MNAS – Modelling Natural Action Selection*. Edinburgh, Scotland 2005.
[8] Brom, C., Abonyi A.: Petri-Nets for Game Plot. In: *Proceedings of AISB – Artificial Intelligence and Simulation Behaviour Convention*, Bristol (2006) **III**, 6–13
[9] Champandard, A. J.: *AI Game Development: Synthetic Creatures with learning and Reactive Behaviors*. USA: New Riders 2003.
[10] Gibson, J. J.: *The Ecological Approach to Visual Perception*. Boston: Houghton Muffin 1979.
[11] Projekt IVE, webovská stránka: http://urtax.ms.mff. cuni.cz/ive/public/about.php
[12] Kubr, J., Šerý, O., Poch, T., Šafrata, P., Kulhánek, J., Šulc, Z.: Dokumentace projektu IVE, MFF UK 2006.
[13] Pešková K.: *Model paměti pro animata.* Online stránka k bakalářské práci. http://carolina.mff.cuni.cz/~peskova/bakalarka/ [2.1.2007]
[14] Šerý, O., Poch, T., Šafrata, P., Brom, C.: Level-Of-Detail in Behaviour of Virtual Humans. In: *Proceedings of SOFSEM 2006: Theory and Practice of Computer Science. LNCS* **3831** (2006) 565–574.
[15] Russell S. J., Norvig P.: *Artificial Intelligence, A Modern Approach.* Englewood Cliffs, New Jersey: Prentice Hall 1995.
[16] Searle, J. R.: Collective intentions and actions. In: *Intentions in Communication*. Cambridge, MA: MIT Press 1990, pp. 401–416.
[17] Stoffregen, T. A.: Affordances as Properties of the Animal-Environment System. *Ecological Psychology* **15**/2 (2003) 115–134.
[18] Tambe, M.: Towards flexible reamwork. *Journal of AI Research* **7** (1997) 203–242.
[19] Turvey, M. T.: Affordances and prospective control: An outline of the ontology. *Ecological Psychology* **4** (1992) 173–187.
[20] Willson, S. W.: The Animat Path to AI. In: Meyer, J.-A., Wilson, S. W.: *From animals to animats*. Cambridge, MA: MIT Press 1991, pp.15–21.
[21] Wooldridge, M.: *An Introduction to MultiAgent Systems*. Chichester: John Wiley & Sons 2002.

Učení neuronových sítí se schopností generalizace

Věra Kůrková*

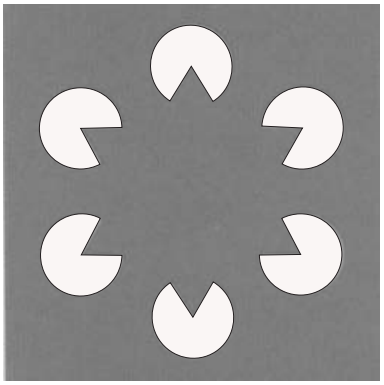
1 Úvod

Metaforicky můžeme strojům říkat umělé svaly, protože nahrazují a rozšiřují schopnosti našich svalů. Podobně dalekohledy, mikroskopy, rentgeny, tomografy a mnohé další měřící přístroje můžeme považovat za umělé smysly mnohonásobně zvyšující schopnosti našich přirozených smyslů. Často ale naměřené údaje, které nám tyto umělé smysly poskytují, nebývají pro nás srozumitelné. Obklopeni množstvím dat, kterým nerozumíme, jsme jako král Midas, který si teprve poté, co se splnilo jeho pošetilé přání, aby se vše, čeho se dotkne, proměnilo ve zlato, uvědomil, že se zlata nenají. K umělým svalům a smyslům potřebujeme ještě umělé neurony, které dovedou naměřená data překládat do nám srozumitelných tvarů. Podobně jako k vynálezům jednoduchých strojů vedlo porozumění zákonům působení síly (byly známy již Archimédovi ve 3. století př. n. l. [26]) a k vynálezům optických přístrojů v 17. století vedlo rozvinutí optiky, pro rozvoj umělých neuronů je třeba porozumění principům analýzy dat.

Jak ale zdůrazňoval Popper [25], z pouhých empirických dat nelze odvodit žádné tvary, pokud o těchto tvarech nemáme předem nějaké *hypotézy*. I naše vizuální vnímání se řídí hypotézou, že se díváme na pravidelné tvary. Například na obr. 1 spatřujeme hvězdu, ačkoliv by se zobrazený nesouvislý tvar dal doplnit mnoha jinými méně pravidelnými tvary. Mnohé úlohy v testech IQ požadující doplnění číselných nebo grafických tvarů jsou rovněž založeny na předpokladu, že hledané tvary vyhovují určité hypotéze, kterou lze nahlédnout na základě několika příkladů. Také v historii vědy nalezneme příklady hledání tvarů, které nejlépe odpovídají naměřeným datům, mezi předem danými tvary vyhovujícími určité hypotéze. Například Kepler předpokládal, že se planety pohybují po nejdokonalejších křivkách, které vyhovují výsledkům měření Tycho Brahe. Když zjistil, že Tychova data neodpovídají kružnicím, zkusil elipsy a tak objevil zákony pohybu planet.

Také Gauss a Legendre na počátku 19. století hledali křivky odpovídající astronomickým datům, tentokrát datům získaným při pozorování komet. Pro nalezení parametrů vhodné křivky použili *metodu nejmenších čtverců*, která hledá funkci s minimálním součtem čtverců chyb na daných empirických datech. Legendre napsal v r. 1805 v Nouvelle

*Ústav informatiky, Akademie věd České republiky, Pod Vodárenskou věží 2, 182 07, Praha 8, E-mail: vera@cs.cas.cz



Obrázek 1.

méthodes pour la determination des orbit des cometes „myslím, že ze všech principů, které mohou být navrženy, není žádný více obecný, přesný a snadněji aplikovatelný než ten, který spočívá v minimalizaci součtu čtverců chyb". Gauss navíc ukázal, že metoda nejmenších čtverců má výhodné statistické vlastnosti a tak se tato metoda stala základem klasických statistických metod zpracování dat [5].

Statistické metody jsou založeny na hypotéze, že data lze modelovat pomocí funkcí, které závisejí na parametrech *lineárně*. Předpoklad, že empirická data odpovídají nějaké lineární závislosti je ale značně omezující. Zejména není vhodný pro vysoce dimenzionální data. Z teorie lineární aproximace je totiž známo, že dimenze lineárního prostoru potřebná pro to, aby se zajistila aproximace hladkých funkcí d proměnných s přesností ε, je $\mathcal{O}\left((\frac{1}{\varepsilon})^d\right)$ [22]. Složitost lineárních modelů potřebná pro zajištění dostatečně přesné aproximace roste tedy exponenciálně s dimenzí dat d.

V ranném období kybernetiky ve 40. letech 20. století se objevily úvahy o *umělých neuronových sítích*, které vedly v 90. letech k vytvoření *konekcionistických výpočetních systémů* schopných klasifikace a rozpoznávání tvarů. Na rozdíl od klasické umělé inteligence, která modelovala kognitivní úlohy jako manipulace se symboly založené na pravidlech, konekcionistické systémy nahrazují namáhavé hledání pravidel *učením na základě příkladů*. Tyto systémy byly inspirovány zjednodušenými představami o mozku tvořeném sítěmi vzájemně propojených neuronů s měnícími se propustnostmi jednotlivých propojení. Formálně je lze popsat jako *transformace vstupních kódů na výstupní kódy* (např. systém NETtalk, který „čte nahlas" anglický text, transformuje grafický kód na fonetický, zatímco systém, který prověřuje žádosti o úvěr, transformuje kódovanou informaci o žadateli na kód vyjadřující rizikovost úvěru). Flexibilita konekcionistických systémů je dána tím, že obsahují mnoho proměnných *parametrů* (kterým se někdy říká váhy), jejichž nastavení určuje transformaci vstupních dat na výstupní data (funkci vstup-výstup), kterou systém provádí.

Umělé neuronové sítě značně rozšířily možnosti zpracování empirických dat. Využívají totiž mnohem bohatší množiny hypotetických funkcí, mezi nimiž lze najít funkce lépe odpovídající datům než v lineárních modelech. Původně byly výpočetní jednotky, z nichž jsou sítě sestaveny, inspirovány biologicky, později se začaly užívat další typy výpočetních jednotek s vhodnými matematickými vlastnostmi pro aproximaci funkcí mnoha proměn-

ných.

Neuronové sítě také přinesly do datové analýzy novou terminologii - hledání parametrů křivek odpovídajících naměřeným datům se začalo říkat *učení na základě příkladů*, vzorkům dat *trénovací množiny* a schopnosti správně zpracovat i data, která nebyla použita pro učení, *generalizace*.

V tomto článku ukazujeme, že pro matematické modelování generalizace lze využít pojmy, které byly vyvinuty pro řešení fyzikálních problémů. Schopnost generalizace při učení na základě příkladů totiž hraje podobnou roli jako *stabilita* při hledáni řešení tzv. *inverzních úloh*. Využití matematických výsledků z teorie inverzních úloh vede k modelům neuronových sítí s jinými typy jednotek než jsou původní biologicky motivované perceptrony. Tyto sítě obsahují tzv. jádrové jednotky. Jádra, která byla studována v matematice po několik staletí v souvislosti s řešením integrálních rovnic, mohou transformovat geometrii vstupních dat tak, aby se data určená ke klasifikaci dala lépe separovat. Kromě toho poskytují matematický popis mnoha typů oscilací a tak umožňují vyloučení nežádoucích řešení. Na základě výsledků teorie inverzních úloh aplikovaných na prostory funkcí definované pomocí jader lze navrhnout algoritmy učení založené na řešení soustav lineárních rovnic.

2 Učení neuronových sítí

Většina algoritmů učení neuronových sítí se snaží nastavit parametry sítě tak, aby síť počítala takovou funkci zobrazující vstupní data na výstupní (říká se jí *funkce vstup-výstup*), která správně transformuje data z trénovací množiny. Pro běžné modely neuronových sítí je ale takových funkcí velké množství. Je třeba mezi nimi vybrat funkci, která správně zobrazuje i data, která nebyla použita při učení sítě. O síti s parametry určujícími takovou funkci se říká, že má schopnost *generalizace*.

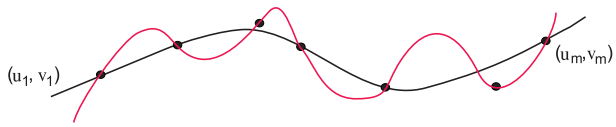
Počátkem 20. století se stal pojem generalizace předmětem filozofického zkoumání. Husserl zavedl pojem *eidetické generalizace* jako procesu představování si možných případů místo pozorování skutečných. Vlastnosti ideálních entit nazýval *eidos* a možné změny entit, které nevedou ke ztrátě eidos, nazýval *eidetické variace*. Eidos lze získat nalézáním podstatných vlastností individuí daného typu. Husserl ale varoval, že takové nalézání zůstává vždy neúplné [28].

Podobně i neuronové sítě se učí rozpoznávat data daného typu na základě příkladů. Algoritmy učení mění parametry sítě tak, aby se snížila hodnota funkcionálu empirické chyby určeného trénovací množinou [32]. Trénovací množina dvojic dat vstup-výstup je v nejjednodušším případě matematicky popsána vektorem $z = \{(u_i, v_i) \in \mathbb{R}^d \times \mathbb{R} \,|\, i = 1, \ldots, m\}$, kde (u_i, v_i) jsou příklady dvojic vstup-výstup, m je velikost vzorku dat, $\mathbb{R}$ značí množinu reálných čísel a d je dimenze vstupních vektorů sítě. *Funkcionál empirické chyby* $\mathcal{E}_z$ určený touto trénovací množinou počítá průměr součtu čtverců chyb, tj. pro funkci f

nabývá hodnoty

$$\mathcal{E}_z(f) = \frac{1}{m}\sum_{i=1}^{m}(f(u_i) - v_i)^2.$$

Pokud je cílem učení nalezení funkce vstup-výstup f, pro niž je hodnota chybového funkcionálu $\mathcal{E}_z(f)$ dostatečně malá, neuronová síť správně zobrazuje data z trénovací množiny, ale není zaručena její schopnost dobře zpracovávat i data, která nebyla použita při učení. Jak zdůrazňoval Popper, z pouhých empirických dat bez předem daných hypotéz nelze odvodit žádné tvary. Jediná informace, kterou v sobě obsahuje funkcionál empirické chyby je informace o vzorku správně zpracovaných dvojic dat $z = \{(u_i, v_i) \mid i = 1, \ldots, m\}$. Hypotéza o hledané funkci je však automaticky obsažena ve volbě množiny funkcí vstup-výstup, v níž algoritmus hledá funkci, pro niž je hodnota funkcionálu $\mathcal{E}_z$ malá (obr. 2). Tato množina je určena volbou typu neuronové sítě a mezí velikosti jejích parametrů. Pro mnohé typy sítí jsou však tyto množiny příliš velké a nepředstavují žádné významné omezení na typ hledané funkce.



Obrázek 2.

Poggio a Girosi [23] počátkem 90. let 20. století nahradili minimalizaci funkcionálu empirické chyby $\mathcal{E}_z$ minimalizací funkcionálu $\mathcal{E}_z + \gamma\Psi$ definovaného

$$\mathcal{E}_z(f) + \gamma\Psi(f) = \frac{1}{m}\sum_{i=1}^{m}(f(u_i) - v_i)^2 + \gamma\Psi(f),$$

kde Ψ je funkcionál nazývaný *stabilizátor* penalizující nežádoucí řešení. Girosi a Poggio použili jako stabilizátor vysokofrekvenční filtr

$$\Psi(f) = \frac{1}{(2\pi)^{d/2}}\int_{\mathcal{R}^d}\frac{\tilde{f}(\omega)^2}{\tilde{k}(\omega)}d\omega, \tag{1}$$

s funkcí k, která má pozitivní Fourierovou transformaci.

Při minimalizaci funkcionálu $\mathcal{E}_z + \gamma\Psi$ se hledá funkce, která nejen že dobře aproximuje vzorek dat $z = \{(u_i, v_i) \in \mathbb{R}^d \times \mathbb{R}, \mid i = 1, \ldots, m\}$, ale také má pokud možno co nejmenší oscilace vyšších frekvencí. Volba stabilizátoru tedy vyjadřuje hypotézu o tvaru řešení, která hraje roli *konceptuálních dat*, která je třeba přidat k *empirickým datům* z, abychom mohli zajistit schopnost generalizace při učení.

3 Inverzní úlohy

Úloha nalézt funkci aproximující empirická data patří do širší třídy tzv. *inverzních problémů*, které se zabývají hledáním *neznámých příčin* (např. sil, tvarů funkcí či distribucí) *zná-*

mých následků (naměřených dat). Řešení inverzních problémů je základním úkolem mnoha oblastí aplikovaného výzkumu jako např. lékařská diagnostika (tomografie), seismologie a meteorologie. Závislost následků na příčinách je zpravidla popsána ve tvaru operátoru, v nejjednodušším případě lineárního, přiřazujícího příčinám (nazývaným *řešení*) následky (nazývané *data*).

Inverzní problém určený operátorem $A : (X, \|.\|_X) \to (Y, \|.\|_Y)$ mezi Banachovými prostory je úloha pro dané $g \in Y$ najít $f \in X$ takové, že

$$A(f) = g.$$

Prvky X se nazývají *řešení* a prvky Y *data*. Je-li Y konečně dimenzionální, inverzní problém se nazývá problém s *diskrétními daty*.

Hadamard v r. 1902 ukázal, že některé klasické fyzikální problémy mají pro všechna data z daného prostoru jediné řešení závisející spojitě na datech. Takové problémy nazval *korektní* (angl. well-posed). Domníval se, že problémy, které jsou nekorektní (tj. problémy, které nemají pro všechna data řešení nebo řešení není jednoznačné a nebo nezávisí spojitě na datech), nemají fyzikální smysl. Avšak fyzika, kterou se zabýval Hadamard, byla fyzika 19. století směřující k Laplaceovu ideálu jednoznačného, stabilního a deterministického popisu světa. V 19. století byly nekorektní problémy považovány za anomálie podobně jako spojité funkce bez derivace, Peanova křivka, Cantorovo diskontinuum a ďábelské schodiště, které vedly ve druhé polovině 20. století k vytvoření fraktální geometrie.

Vývoj od anomálie k centru zájmu aplikované matematiky proběhl také v oblasti inverzních úloh. Jedna z nejúspěšnějších aplikací řešení nekorektní úlohy je počítačová tomografie založená na výpočtu inverzní Radonovy transformace. Za její vynález byla v r. 1979 udělena Nobelova cena za medicínu [8]. Počítačový obraz plošného řezu části těla je řešením inverzní úlohy s daty získanými rentgenováním podél velkého počtu přímek ze zobrazované plochy. Matematicky se jedná o rekonstrukci funkce definované na ploše z integrálů podél přímek ležících v této ploše.

Metodologie řešení nekorektních úloh je založená na *Mooreově-Penroseově pseudoinverzi* a různých typech *regularizace*. Pro konečně dimenzionální problémy, které nemají pro všechna data řešení, navrhl v r. 1920 Moore [18] metodu zobecněné inverze založenou na hledání tzv. *pseudořešení*. Jeho práce publikovaná pouze jako abstrakt zůstala nedoceněna a metoda pseudoinverze nebyla dále rozvíjena ani využívána až do jejího znovuobjevení Penrosem v r. 1955 [21]. I když řešení neexistuje, lze hledat alespoň jeho nejlepší aproximaci nazývanou *pseudořešení* f^o splňující rovnici

$$\|A(f^o) - g\|_Y = \min_{f \in X} \|A(f) - g\|_Y$$

a v případě, že je těchto pseudořešení více, hledat mezi nimi *normální pseudořešení* f^+, které má nejmenší normu. Pokud pro všechna $g \in Y$ existuje normální pseudořešení f^+, pak může být definován *pseudoinverzní operátor* $A^+ : Y \to X$ jako

$$A^+(g) = f^+.$$

Pro X a Y konečně dimenzionální, kdy lze lineární operátor A popsat maticí, pseudoinverzní operátor odpovídá Moore-Penroseově pseudoinverzi matice. Teorie pseudoinverze byla rozšířena na lineární operátory mezi Hilbertovými prostory [11].

Vzhledem k tomu, že empirická data podléhají nepřesnostem měření, podstatnou vlastností operátoru popisujícího inverzní problém je stabilita vzhledem k šumu. Pokud vede malý rozdíl v datech způsobený chybou měření k velké změně řešení, je takové řešení nestabilní. Je-li A spojitý, pak podle Banachovy věty [9, p.141] je A^{-1} také spojitý. Ale ani spojitá závislost řešení na datech nezaručí vždy stabilitu vzhledem k šumu. Jako míra stability řešení inverzního problému se používá tzv. *číslo podmíněnosti* definované pro korektní problém jako $cond(A) = \|A\| \, \|A^{-1}\|$ a pro nekorektní jako $cond(A) = \|A\| \, \|A^{+}\|$. Dá se snadno ukázat, že je-li $A(f) = g$ a $A(f') = g'$, potom $\frac{\|f-f'\|_X}{\|f\|_X} \leq cond(A) \frac{\|g-g'\|_Y}{\|g\|_Y}$. Relativní odchylka řešení je tedy menší nebo rovna součinu čísla podmíněnosti a relativní odchylky dat. Číslo podmíněnosti je vždy větší nebo rovno 1. Pokud je malé, problém se nazývá *dobře podmíněný*, pokud je velké, hovoří se o *špatně podmíněném problému*.

Metoda *regularizace* vyvinutá v 60. letech 20. století je založena na myšlence, že se lze zbavit nejednoznačnosti a zlepšit stabilitu řešení, pokud kromě empirických dat vezmeme také v úvahu *konceptuální data* popisující nějakou globální vlastnost hledané funkce. To znamená že omezíme prostor, kde hledáme řešení vyhovující naměřeným datům, pouze na řešení, která mají fyzikální smysl, tj. splňují nějakou předem danou podmínku (jako např. omezený výskyt oscilací vyšších frekvencí) nebo penalizujeme nežádoucí řešení. Metodu objevilo na sobě nezávisle několik autorů, jednotný rámec jí dal Tikhonov [29]. Tikhonovova regularizace nahrazuje problém minimalizace funkcionálu $\|A(.) - g\|_Y^2$ na prostoru X hledáním funkce f^γ splňující

$$\|A(f^\gamma) - g\|_Y^2 + \gamma \|f^\gamma\|_X^2 = \min_{f \in X} \|A(f) - g\|_Y^2 + \gamma \|f\|_X^2.$$

Regularizace tedy penalizuje řešení s velkou normou $\|.\|_X$. Tato norma může modelovat nějakou vlastnost řešení jako například hladkost funkce, její lokalizaci nebo výskyt určitého typu oscilací. Pomocí regularizačního parametru γ lze měnit míru této penalizace.

4 Učení jako inverzní úloha

Minimalizaci funkcionálu empirické chyby lze vyjádřit jako inverzní úlohu popsanou *evaluačním operátorem* $L_u : X \to \mathbb{R}^m$ definovaným

$$L_u(f) = \left(\frac{f(u_1)}{\sqrt{m}}, \ldots, \frac{f(u_m)}{\sqrt{m}} \right).$$

Protože funkcionál empirické chyby $\mathcal{E}_z$ lze vyjádřit jako

$$\mathcal{E}_z = \left\| L_u - \frac{v}{\sqrt{m}} \right\|_2^2,$$

kde $\|.\|_2$ značí l_2-normu na $\mathbb{R}^m$, problém minimalizace $\mathcal{E}_z$ na X je ekvivalentní úloze najít pseudořešení inverzního problému daného operátorem L_u pro data $\frac{v}{\sqrt{m}}$.

Zobecnění Moore-Penroseovy teorie [4, pp. 56-60], [11, pp. 37-46] dobře popisuje vlastnosti řešení a jeho stability v případě, že operátor je spojitý a má uzavřený obraz. Protože obraz operátoru L_u je konečně dimenzionální, je uzavřený. Ale na rozdíl od mnoha inverzních problémů motivovaných fyzikálními ději operátor L_u není spojitý vzhledem k $\mathcal{L}_2$-normě (dá se ukázat, že není omezený [15]). Nemůžeme tedy odvodit výsledky pro minimalizaci na celém prostoru všech $\mathcal{L}_2$-funkcí.

Pokud vyloučíme funkce, které mají příliš mnoho vysokofrekvenčních oscilací, přidáme tak k empirickým datům také data konceptuální ve formě podmínky omezující určitý typ oscilací. Formálně lze globální vlastnosti takových řešení modelovat tak, že za obor hodnot operátoru L_u zvolíme Hilbertův prostor obsahující jen funkce splňující určitá omezení na vhodný typ oscilací. Prostory tohoto typu se nazývají *Hilbertovy prostory s reprodukčním jádrem* (anglicky reproducing kernel Hilbert spaces a proto se označují akronymem RKHS). Kromě toho, že se RKHS velice dobře hodí pro modelování konceptuálních vlastností vhodných pro matematickou formalizaci schopnosti generalizace, jsou na nich operátory L_u spojité [2].

RKHS jsou jednoznačně určeny *jádry*, tj. *symetrickými pozitivně semidefinitními funkcemi* $K : \Omega \times \Omega \to \mathbb{R}$, kde Ω je libovolná neprázdná množina. Výklad teorie těchto prostorů přesahuje rámec této práce, jsou podrobně popsány v knihách [3], [31], stručnější přehledy jejich vlastností lze nalézt např. v [7]. Pro ilustraci, jakou roli hrají normy na těchto prostorech, uvedeme jen následující příklad: Je-li K konvoluční jádro, tj. K lze vyjádřit jako $K(x, y) = k(x - y)$ kde $k : \mathbb{R} \to \mathbb{R}$, a jestliže k má pozitivní Fourierovu transformaci, potom norma na RKHS určeném K označovaná $\|.\|_K$ má tvar

$$\|f\|_K^2 = \frac{1}{(2\,\pi)^{d/2}} \int_{\mathbb{R}^d} \frac{\tilde{f}(\omega)^2}{\tilde{k}(\omega)}\, d\omega.$$

To znamená, že pro konvoluční jádra s pozitivní Fourierovou transformací tvoří normy na RKHS vysokofrekvenční filtry. Paradigmatický příklad jádra je konvoluční jádro definované pomocí Gaussovy funkce $\exp(-b\|x\|^2)$, jejíž Fourierova transformace je $(2b)^{-d/2}$ $\exp(-\|x\|^2/4b)$.

S využitím teorie pseudoinverze aplikované na operátory L_u na RKHS prostorech se dají elegantně odvodit vlastnosti řešení úlohy minimalizace funkcionálu empirické chyby $\mathcal{E}_z$ na RKHS. Pro jednoduchost popíšeme jen případ, kdy jádro splňuje silnější podmínku pozitivní definitnosti místo semidefinitnosti [3]. Pro tato jádra je řešení tvaru

$$f^+(x) = \sum_{i=1}^{m} c_i K(u_i, x),$$

kde vektor koeficientů $c = (c_1, \ldots, c_m)$ splňuje $c = \mathcal{K}[u]^+ v$ ($\mathcal{K}[u]$ je Gramova matice jádra K vzhledem k vektoru vstupních dat $u = (u_1, \ldots, u_m)$ definovaná jako $\mathcal{K}[u]_{i,j} = K(u_i, u_j)$, $i, j = 1, \ldots, m$) [13]. To znamená, že řešení se dá spočítat pomocí Moore-Penrosovy pseudoinverze matice $\mathcal{K}[u]$ aplikované na vektor výstupních dat $v = (v_1, \ldots, v_m)$. Toto řešení

interpoluje data (minimum $\mathcal{E}_z$ je rovno nule) a jeho stabilita vzhledem k malé změně vektoru v závisí na čísle podmíněnosti matice $\mathcal{K}[u]$.

5 Generalizace modelovaná pomocí regularizace

Globální podmínka modelující schopnost generalizace může být ještě zesílena tím, že nejen hledáme řešení pouze mezi funkcemi s konečnou K-normou, ale navíc penalizujeme řešení, která mají tuto normu velkou. V případě problému učení na základě empirických dat minimalizace hodnoty

$$\mathcal{E}_z(f) + \gamma\|f\|_K^2 = \|L_u(f) - \frac{v}{\sqrt{m}}\|_2^2 + \gamma\|f\|_K^2$$

modeluje úlohu hledání funkce, která nemá příliš velké oscilace typu daného volbou jádra K a současně dobře aproximuje vzorek dat z. Dá se ukázat, že tato úloha má jediné řešení tvaru

$$f^\gamma(x) = \sum_{i=1}^{m} c_i^\gamma K(u_i, x),$$

kde $c^\gamma = (\mathcal{K}[u] + \gamma m\mathcal{I})^{-1}v$ a $\mathcal{I}$ značí identickou matici [7], [24], [13].

V regularizovaném i neregularizovaném případě je tedy ideálním řešením úlohy učení neuronová síť, která má stejný počet výpočetních jednotek m ve skryté vrstvě jako je velikost trénovací množiny a tyto jednotky počítají jádrové funkce $K(u_i, .)$ s parametry u_i danými vstupními daty (v případě radiálních sítí se těmto parametrům říká centroidy). Tato dvě řešení se liší pouze koeficienty lineární kombinace výpočetních jednotek, které odpovídají výstupním vahám sítě. V neregularizovaném případě platí $c = \mathcal{K}[u]^+v$, vektor výstupních vah c je tedy roven Moore-Penrose pseudoinverzi Gramovy matice jádra K vzhledem k vektoru vstupních dat u aplikované na vektor výstupních dat v. V regularizovaném případě platí $c^\gamma = (\mathcal{K}[u] + \gamma m\mathcal{I})^{-1}v$, vektor výstupních vah c^γ je tedy roven inverzi matice $\mathcal{K}[u] + \gamma m\mathcal{I}$ aplikované na vektor v.

Vzhledem k tomu, že obě matice jsou pozitivně definitní, lze vyjádřit jejich čísla podmíněnosti pomocí vlastních hodnot a na základě tohoto vyjádření odvodit vztah

$$cond(\mathcal{K}[u] + \gamma m\mathcal{I}) = 1 + \frac{(cond(\mathcal{K}[u]) - 1)\lambda_{\min}}{\lambda_{\min} + \gamma m},$$

kde $\lambda_{\min}$ značí nejmenší vlastní hodnotu matice $\mathcal{K}[u]$. Zvyšováním regularizačního parametru γ se tedy zlepšuje stabilita řešení. Tato výhoda regularizace je však omezená nutností zachovat dobrou aproximaci empirických dat, která se zhoršuje se zvyšováním penalizace za oscilující řešení.

Pokud nejsou čísla podmíněnosti matic $\mathcal{K}[u]$ nebo $\mathcal{K}[u] + \gamma m\mathcal{I}$ příliš velká, lze využít charakterizace řešení problému učení odvozené pomocí aplikace teorie pseudoinverzních operátorů pro návrh algoritmů učení [24]. Pro mnohá jádra a vzorky dat mají ale Gramovy matice velká čísla podmíněnosti [19]. V takovém případě vedou tyto algoritmy k řešením,

která jsou nestabilní vzhledem k šumu. Na rozdíl od těchto algoritmů založených na řešení soustav lineárních rovnic, algoritmy využívající zpětné šíření chyb nebo evoluční principy hledají suboptimální řešení dosažitelná sítěmi s počtem prvků n mnohem menším než je velikost trénovací množiny m. V některých případech tato suboptimální řešení dobře aproximují výše popsaná optimální řešení. Odhady rychlosti jejich konvergence s rostoucím počtem prvků byly odvozeny v [16, 17] na základě metod nelineární aproximace popsaných v [12].

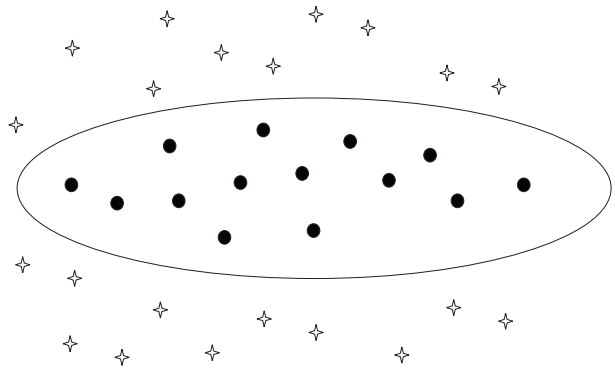
6 Jádrové metody

Koncem 90. let Girosi [10] zjistil, že stabilizátor (1) je speciálním příkladem čtverce normy na tzv. *Hilbertově prostoru s reprodukčním jádrem*. Tyto prostory označované akronymem *RKHS (reproducing kernel Hilbert spaces)* definoval v r. 1950 Aronszajn [2], ale jejich teorie zahrnuje mnoho klasických matematických poznatků týkajících se integrálních rovnic s jádry a pozitivně definitních funkcí [3]. Pomocí norem na těchto prostorech se dají formálně popsat nejrůznější typy oscilací. Od Aronszajnovy definice k prvním aplikacím v analýze dat ale uplynula dost dlouhá doba, až v 60. letech je využil Parzen [20] a později Wahba [31] pro aproximaci dat pomocí splinů.

Využití norem na RKHS jako stabilizátorů při regularizaci funkcionálu empirické chyby tedy umožňuje stanovit různé apriorní podmínky na hledanou funkci aproximující data z trénovací množiny. V případě stabilizátoru (1) se jedná o čtverec normy na RKHS definovaném pomocí tzv. konvolučního jádra $K(x, y) = k(x - y)$, kde $k > 0$.

Jádra, která definují RKHS, ale našla uplatnění v teorii učení i ze zcela jiného důvodu, než že se normy, které tato jádra určují, dobře hodí jako míry různých typů oscilací. V 60. letech Aizerman, Braverman a Rozonoer [1] využili jádra (která nazývali potenciálové funkce) pro modifikaci geometrie prostoru dat. Pomocí vhodně zvoleného jádra lze totiž definovat vnoření prostoru vstupních dat $\mathbb{R}^d$ do RKHS tak, že se dvě množiny dat, které se v $\mathbb{R}^d$ nedají lineárně oddělit, zobrazí na lineárně separovatelné množiny (obr. 3). V RKHS s takovým jádrem lze pak snadno provést klasifikaci pomocí jednoho prahového prvku (perceptronu). Algoritmus navržený Aizermanem, Bravermanem a Rozonoerem [1] rozvinul Vapnik [30], který nazval výpočetní model provádějící klasifikaci pomocí jader *support vector machine (SVM)*. V terminologii neuronových sítí je SVM síť s jednou skrytou vrstvou tvořenou jádrovými jednotkami a s jednou prahovou výstupní jednotkou. V současné době jsou jádrové metody předmětem intenzivního výzkumu jak v oblasti teorie tak v oblasti aplikací (viz např. monografie zaměřené na aplikace [27] a [6], rozsáhlý teoretický článek [7] nebo stručný přehled [24]).

Jádra jsou tedy vhodnými prvky výpočetních modelů ze dvou důvodů: umožňují měnit skalární součin a tím geometrii prostorů dat a vyjádřit míru nežádoucích oscilací funkcí vstup-výstup. Dalším důvodem pro využití jader v teorii učení je *spojitost evaluačních funkcionálů* na prostorech RKHS definovaných pomocí jader. Je to dokonce jedna z jejich dvou ekvivalentních definic [2] (ta druhá stanoví, že je prostor definován pomocí pozitivně semidefinitního jádra [3]).



Obrázek 3.

7 Závěr

Metody řešení matematických problémů z oblasti učení neuronových sítí zapadají do dlouhodobého vývoje matematických idejí původně motivovaných fyzikálními úlohami. Pomocí vhodně definovaného lineárního operátoru lze formulovat úlohu učení na základě empirických dat jako inverzní problém. Tato reprezentace umožňuje popis optimálního řešení a modelování schopnosti generalizace. Také vede k odhadům zlepšení stability řešení přidáním konceptuálních dat vhodného tvaru.

Poděkování: Tato práce vznikla s částečnou podporou grantu GA ČR 201/05/0557 a Institucionálního záměru AV0Z10300504.

Literatura

[1] Aizerman, M. A., Braverman, E. M., Rozonoer, L. I.: Theoretical foundations of potential function method in pattern recognition learning. *Automation and Remote Control* **28** (1964) 821–837.

[2] Aronszajn, N.: Theory of reproducing kernels. *Transactions of AMS* **68** (1950) 33 – 404.

[3] Berg, C., Christensen, J. P. R., Ressel, P.: *Harmonic Analysis on Semigroups*. New York: Springer-Verlag 1984.

[4] Bertero, M.: Linear inverse and ill-posed problems. *Advances in Electronics and Electron Physics* **75** (1989) 1 – 120.

[5] Bjorck, A.: *Numerical methods for least squares problem*. SIAM 1996.

[6] Cristianini N., Shawe-Taylor J.: *An Introduction to Support Vector Machines*. Cambridge: Cambridge University Press 2000.

[7] Cucker, F., Smale, S.: On the mathematical foundations of learning. *Bulletin of AMS* **39** (2002) 1 – 49.

[8] Di Chiro, G., Brooks, R. A. : The 1979 Nobel prize in physiology or medicine, *Science* **206** (1979) 1060–1062.

[9] Friedman, A.: *Modern Analysis*. New York: Dover 1982.
[10] Girosi, F.: An equivalence between sparse approximation and support vector machines. *Neural Computation* **10** (1998) 1455 – 1480 (AI Memo No 1606, MIT).
[11] Groetch, C. W.: *Generalized Inverses of Linear Operators*. Dekker, New York 1977.
[12] Kurková, V.: High-dimensional approximation by neural networks. Chapter 4 in *Advances in Learning Theory: Methods, Models and Applications* (J. Stuykens et al., Ed.) 69 – 88. Amsterdam: IOS Press 2003.
[13] Kurková, V.: Learning from data as an inverse problem. In *COMPSTAT 2004 - Proceedings on Computational Statistics* (J. Antoch, Ed.) 1377–1384. Heidelberg: Physica-Verlag 2004.
[14] Kurková, V: Neural network learning as an inverse problem. *Logic Journal of IGPL* **13** (2005) 551–559.
[15] Kurková, V: Generalization in learning from examples. In *Challenges for Computational Intelligence*. (Eds. W. Duch, J. Mandziuk) Berlin, Heidelberg: Springer-Verlag, 2007.
[16] Kurková, V., Sanguineti, M.: Error estimates for approximate optimization by the extended Ritz method. *SIAM J. on Optimization* **15** (2005) 461–487.
[17] Kurková, V., Sanguineti, M.: Learning with generalization capability by kernel methods with bounded complexity. *Journal of Komplexity* **21** (2005) 350–367.
[18] Moore, E. H.: Abstract. *Bulletin AMS* **26** (1920) 394–395.
[19] Narcowich, F. J., Sivakumar, N., Ward, J. D. On condition numbers associated with radial-function interpolation. *Journal of Mathematical Analysis and Applications* **186** (1994) 457 – 485.
[20] Parzen, E.: An approach to time series analysis. *Annals of Math. Statistics* **32** (1966) 951 – 989.
[21] Penrose, R.: A generalized inverse for matrices. *Proceedings of Cambridge Philos. Soc.* **52** (1955) 406–413.
[22] Pinkus, A.: *n-width in Approximation Theory.* Berlin: Springer-Verlag 1985.
[23] Poggio, T., Girosi, F.: Networks for approximation and learning. *Proceedings of IEEE* **78** (1990) 1481 – 1497.
[24] Poggio, T., Smale, S.: The mathematics of learning: dealing with data. *Notices of the AMS* **50** (2003) 536 – 544.
[25] Popper, K.: *The Logic of Scientific Discovery*. New York: Harper Torch Book 1968.
[26] Russo, L.: *The Forgotten Revolution*. Berlin: Springer-Verlag, 2004.
[27] Schölkopf B., Smola A. J.: *Learning with Kernels- Support Vector Machines, Regularization, Optimization and Beyond*. Cambridge: MIT Press 2002.
[28] Smith, D. W., McIntyre, R.: *Husserl and Intentionality: A Study of Mind, Meaning, and Language*. Dordrecht, Boston: D. Reidel Publishing Co. 1982.
[29] Tikhonov, A. N., Arsenin, V. Y.: *Solutions of Ill-posed Problems*. Washington, D.C.: W. H. Winston 1977.
[30] Vapnik V. N.: *The Nature of Statistical Learning Theory*. New York: Springer-Verlag 1995.
[31] Wahba, G.: *Splines Models for Observational Data*. Philadelphia: SIAM 1990.

[32] Werbos, P. J.: *Backpropagation: Basics and New Developments*. The Handbook of Brain Theory and Neural Networks (Ed. Arbib M.) (pp. 134–139). Cambridge: MIT Press 1995.

Modelovanie na báze reaktívnych agentov

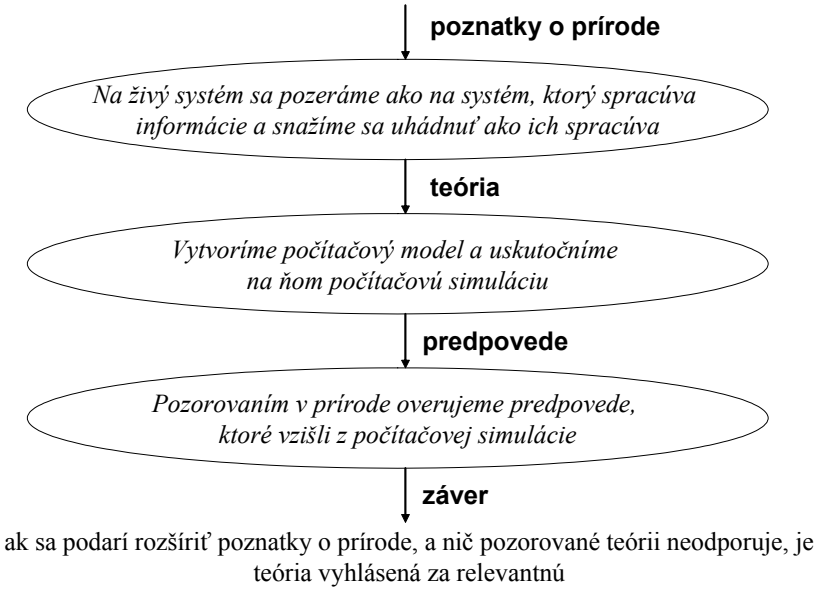
Andrej Lúčny[1]

Abstrakt. Pokiaľ modelujeme pozorované správanie nejakého živého tvora, musíme si dať pozor na to, aby sme nemodelovali samotné pozorovanie, ale štruktúry, ktoré sú zaň v danom tvorovi zodpovedné. Môžeme pritom postupovať tak, že sa snažíme povahu týchto štruktúr uhádnuť a z vytvoreného modelu, ktorý vysvetľuje dané pozorovanie, získať určité testovateľné predpovede. Tento postup demonštrujeme na príklade a uvádzame pritom vlastnú predstavu o povahe týchto štruktúr.

1 Fyzikálny prístup k modelovaniu biologických štruktúr

Pokiaľ pozorujeme správanie nejakého jednoduchého tvora a snažíme sa ho namodelovať počítačovou simuláciou, musíme sa vyvarovať jedného častého omylu. Spočíva v tom, že namiesto toho, aby sme modelovali tvora, ktorý sa pozorovaným spôsobom správa, modelujeme pozorované správanie. Taký model dané správanie ľahko vygeneruje, ale v skutočnosti nepovie vôbec nič, len zaznamená naše pozorovanie. Nás ale nezaujíma ani tak dané správanie ako spôsob, akým povstáva v danom tvorovi z jeho vnútorných štruktúr.

O tom, ako sú tieto štruktúry organizované, máme už z biológie určitú predstavu. V detailoch ju však pre účely počítačovej simulácie musíme značne dotvoriť, čím vlastne hádame aké softwarové štruktúry v počítači sú analogické biologickým štruktúram zodpovedným za generovanie správania.



Obrázok 1. Fyzikálny prístup k modelovaniu biologických štruktúr

[1] Katedra aplikovanej informatiky FMFI UK Bratislava a MicroStep-MIS, www.microstep-mis.com/~andy
E-mail: andy@microstep-mis.com

Táto ucelená predstava potom definuje všeobecné možnosti, v rámci ktorých sa snažíme nájsť také konkrétne štruktúry, ktoré dané správanie generujú. Takýto model však nielen generuje pozorované správanie, ale dokáže poskytnúť množstvo predpovedí pre situácie odlišné od pozorovaných, ktoré sa potenciálne dajú spätne overiť dodatočným pozorovaním. Na základe toho si potom môžeme urobiť aj úsudok o tom, nakoľko kvalitne sme uhádli biologickú relevantnosť štruktúr použitých pri simulácii.

Napodobňujeme tu teda spôsob akým sa vo fyzike skúma vnútorná štruktúra hmoty. Určitý experiment tu vedie k postulovaniu teórie zahrňujúcej určitú predstavu o štruktúre, ktorá ho vysvetľuje. Zároveň ale táto teória generuje určité predpovede, ktoré sa dajú overiť ďalšími experimentmi. Keď sú predpovede potvrdené, posilňuje to presvedčenie, že ide o dobrú teóriu, hoci nikto netvrdí, že je dokonalá (obrázok 1).

Takto získané poznatky potom predstavujú nielen určité biologicky relevantné poznanie, ale aj zdroj inšpirácie pre programovanie tých umelých systémov, ktoré živé systémy do istej miery napodobňujú.

2 Príklad modelovaného správania

Za príklad na demonštrovanie našej myšlienky si zvolíme modelovanie zakladania potomstva kutavkou [4]. Ide o jedno z najzložitejších individuálnych správaní v ríši hmyzu vôbec a zjednodušujúcim faktorom tu je, že toto správanie je evidentne vrodené (žiadna kutavka nevidí svoju matku zakladať potomstvo).

Kutavka je samotárska osa (obrázok 2) známa najmä pre hrôzostrašný spôsob akým zabezpečuje pre svoje larvy dostatok čerstvej potravy: žihadlom pichne svoju korisť (každý druh loví špecifickú obeť z radov chrobákov, napríklad sedlovku) následkom čoho táto ochrnie a odtiahne ju do podzemnej komôrky, kde na ňu nakladie vajíčko. Z toho sa vyliahne larva, ktorá potom paralyzovanú obeť zaživa zje, pričom jej to trvá niekoľko mesiacov. Kutavka pritom realizuje vždy rovnaký rituál[2]: najprv vyhrabe komôrku, potom uloví obeť, položí si ju pred komôrku, vlezie do komôrky skontrolovať ju, potom do nej vtiahne obeť, znesie na ňu vajíčko, vylezie z komôrky a uzavrie jej vchod.

Obrázok 2. Kutavka

S týmto správaním sa dá zaujímavo pohrať. Pokiaľ by ste kutavke vo fáze, keď vlezie na kontrolu do komôrky, sedlovku odtiahli, ale nie príliš ďaleko – aby ju kutavka ľahko našla, opäť by ju priniesla ku komôrke. Ale opäť si ju položí pred vchod a ide vykonať kontrolu (ktorú už práve pred chvíľou vykonala), čiže vlezie do komôrky. Takto má experimentátor možnosť opakovať odtiahnutie sedlovky od vchodu a udržiavať

[2] Tento rituál sa odlišuje v rámci jednotlivých druhov kutaviek, tu uvádzame podobu z [4]

kutavku v bezvýchodiskovom cykle. Kutavka 30-40 krát zakaždým sedlovku pritiahne ku vchodu komôrky, ale nepoučí sa a nevtiahne ju hneď dnu, naopak, ide vykonávať kontrolu, ktorú už vykonala mnoho krát a s pričinením experimentátora opäť o sedlovku príde. E. Gál položil v [3] provokujúcu otázku: čo stojí za týmto, z nášho hľadiska dosť tupým, správaním kutavky?

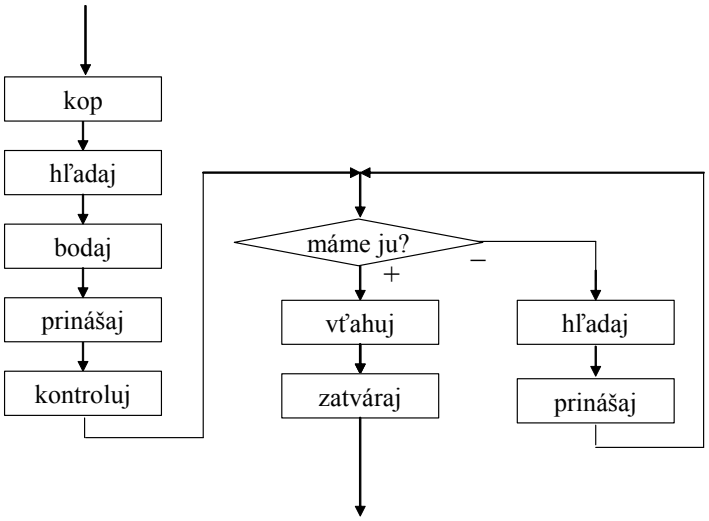
Nejaké jednoduché vysvetlenie tu neobstojí, lebo napríklad na rozdiel od kontroly komôrky, pri opätovnom nájdení odtiahnutej sedlovky, ju kutavka nejde druhý raz pichať žihadlom. Určité fázy vie teda zopakovať, iné preskočiť.

Navyše kutavka si evidentne dokáže zapamätať, že už vykopala komôrku a dokonca kde to bolo. Jej schopnosti učiť sa orientovať pri lete podľa význačných pozemných objektov boli jednoznačne preukázané.

S predstavou, že kutavka je z hľadiska zložitosti svojho riadiaceho mechanizmu len jednooký medzi slepými (teda relatívne jednoduchý tvor vykonávajúci pomerne zložité správanie) sa teraz pokúsime zostrojiť počítačový model uvažovaného správania.

3 Tradičný model

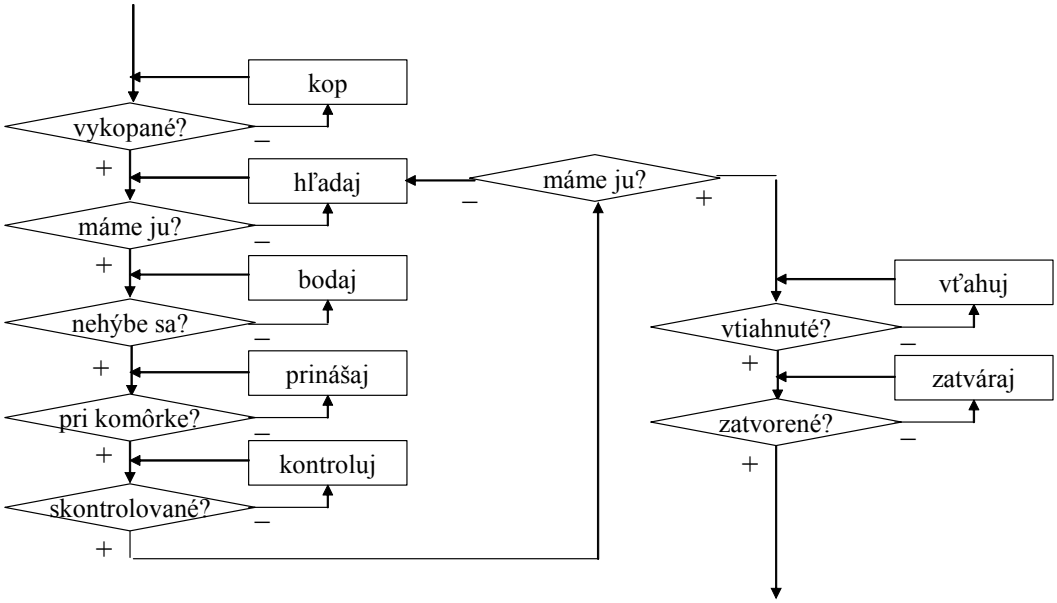
Pokiaľ by sme chceli zvolené správanie zaznamenať algoritmickým spôsobom, zrejme by sme na úvod vytvorili niečo podobné obrázku č. 3. Pokiaľ by sme to naozaj spravili takto, dostali by sme sa k už uvedenému problému zámeny modelovania pozorovania a pozorovaného. Simulátor založený na interpretácii takéhoto algoritmu totiž na jednej strane vykazuje výbornú zhodu so zvoleným pozorovaním, avšak na strane druhej nič iné než pozorované správanie nedokáže vygenerovať.



Obrázok 3. Modelovanie pozorovania (tradičný postup)

Ak potom simulátor čo len trochu vychýlime z podmienok zvoleného experimentu, napríklad počas kontroly komôrky ochrnutú sedlovku odstránime a nahradíme neochrnutou, dostaví sa zlyhanie. Pri prechode od hľadania k prinášaniu sa prinášanie s neochrnutou sedlovkou nepodarí, čo z hľadiska počítačovej simulácie zodpovedá vygenerovaniu výnimky.

Príroda si ale vždy nájde cestu, čo môžeme premietnuť aj do daného algoritmického vyjadrenia modelu. Riešenie na obrázku 4 napríklad predpokladá, že každá výnimka je ošetrená vynechaním danej fázy. Takýto model už dokáže generovať určité predpovede, napríklad, že ak vymeníme sedlovku za neochrnutú, tak ju kutavka bodne, zatiaľ čo pri ponechaní pôvodných podmienok bodnutie vynechá. Potiaž je, že tu ľahko nájdeme aj predpovede načisto nesprávne, napríklad, že keď zakopeme komôrku, tak sa ju do nekonečna snažíme kontrolovať. Dotiahnutie tohto modelu by si vyžadovalo veľké množstvo rôznych podmieňovacích príkazov. Boli by sme samozrejme radšej, keby sme ich vedeli zakódovať do modelu čo najúspornejším spôsobom.



Obrázok 4. Čiastočný prechod k modelovaniu pozorovaného (tradičný postup)

Ďalšie problémy predstavujú časové parametre modelu. Oba uvedené modely napríklad nevysvetľujú ako dôjde k tomu, že kutavka napriek dlhému cykleniu nakoniec pokusy o kontrolu komôrky zanechá. Doplnenie by tu spočívalo v inkrementálnom počítaní „netrpezlivosti“ pri prechodoch cyklami alebo volaním generátora náhody v podmienkach, čo by ešte skomplikovalo celý model.

Zvyšovanie zložitosti modelu sa pritom nepriaznivo odráža nielen v našej schopnosti mu porozumieť, ale aj v znižovaní jeho schopnosti generovať predpovede. Tie totiž zodpovedajú tomu, že explicitným zanášaním určitých vlastností do modelu v skutočnosti pridávame aj niečo čo sme vlastne pridať nezamýšľali, niečo, čo zanášame implicitne. A pri štruktúrach tradičného programovania sa takáto situácia objavuje načisto výnimočne.

4 Alternatívny model

Potrebovali by sme teda do nášho modelu vložiť takú predstavu o štruktúrach generujúcich uvažované správanie, kde jednotlivé jemné odtiene a variácie tohto správania implicitne povstávajú z relatívne malej informácie, explicitne uloženej v týchto štruktúrach.

Samozrejme tutovkou sa tu zdá byť voľba štruktúr veľmi podobných biologickým korelátom ako sú neurónové siete. My si volíme cestu inú a to takú, že sa snažíme odpozorovať len všeobecné vlastnosti biologických štruktúr, presnejšie povahu ich modularity, organizácie a aktivácie. Vedie nás k tomu fakt, že pri tomto prístupe sme stále blízko tradičného programovania, pomocou ktorého – na rozdiel od súčasných neurónových sietí – vieme dosiahnuť jednak škálovateľnosť riešení, jednak jasné rozlíšenie medzi vrodeným a naučeným. (V princípe je ale naša cesta len jedna z mnohých.)

Takže akými dôležitými vlastnosťami sa vyznačujú biologické štruktúry produkujúce správanie? Podľa nás týmito štyrmi:

1. **decentralizovanou modularitou s vysokým paralelizmom.** O paralelnej povahe spracovania informácii v biologických systémoch zrejme nikto nepochybuje. O tom, že z hľadiska riadenia ide o decentralizovaný systém, svedčí rôznorodosť a vysoká špecifickosť porúch, ktoré môžu vplyvom zranenia, alebo dedičnej vady nastať.
2. **nízkou kontinuitou vykonávania inštrukcii.** Na rozdiel od počítača, ktorý dokáže interpretovať rutinu pozostávajúcu z neobmedzene dlhej sekvencie inštrukcii, v prírode je s niečím podobným problém. Pri jednoduchších tvoroch ide z hľadiska určitého správania o jednotky, maximálne desiatky úkonov. Aj medzi ľuďmi sú jedinci disponujúci schopnosťou čo i len zapamätať si dlhú sekvenciu dát nie tak dokonalí, ale tak retardovaní.
3. **obmedzenou časovou platnosťou operačných dát.** Na rozdiel od počítača, kde sú údaje dokonale perzistentné a ich zabúdanie sa musí špecificky implementovať, v prírode je veľa indikácii (vrátane termodynamickej povahy pamäti), že problém je skôr dáta trvalo uchovať.
4. **možnosťou infiltrácie kooperácie dvoch modulov inými modulmi, pričom prevláda potláčanie a korekcia aktivity nad jej vyvolávaním**. Je známych niekoľko experimentov, ktoré svedčia v prospech toho, že hierarchicky vyššie vrstvy skôr potláčajú a usmerňujú aktivitu nižších, než že by ju vyvolávali. Napríklad keď (dostatočne jemne) dekapitujeme pakobylku, tak pekne rovnomerne kráča. Jej nervové uzly v hlave teda uzly v hrudi nestimulujú, ale potláčajú. Nie sú teda s nimi zapojené v sérii, ale hierarchicky narúšajú komunikáciu v uzloch hrude.

Náš alternatívny model bude preto pozostávať z viacerých paralelných jednotiek. Budeme ich nazývať agentami. Ich vzájomná komunikácia bude realizovaná pomocou globálnych premenných, ktoré môže viacero agentov (naraz) zapisovať aj čítať. Každý agent bude vykonávať pomerne krátku a jednoduchú sekvenciu inštrukcií, a to buď pravidelne v čase, alebo na základe zmeny hodnoty niektorej globálnej premennej. Oproti tradičnému chápaniu premenných sa tie naše budú vyznačovať dvomi novinkami:

- hodnota sa do premennej bude dať zapísať aj na obmedzený čas, po ktorom sa stratí
- hodnote budeme môcť definovať prioritu, ktorá spôsobí, že hodnota sa do premennej vôbec nepriradí ak je v nej hodnota s vyššou prioritou

Časovú platnosť i prioritu definuje agent pri priradení hodnoty do premennej. Kým časovú platnosť potrebujeme podľa bodu 3, priority nám umožnia definovať hierarchiu pri napĺňaní bodu 4.

Vyššie uvedené je podstatou nami navrhnutej architektúry agent-space, prednostne určenej na tvorbu riadiacich systémov napr. mobilných robotov. Jej technické detaily sú pomerne zložité a čitateľ ich môže nájsť v [8]. Na pochopenie myšlienky tohto článku však nie sú nevyhnutné.

Z filozofického hľadiska každý agent predstavuje v systéme určitú zabudovanú reakciu, určitý reflex. Možno teda povedať, že tu ide o snahu **vymodelovať správanie systému ako kolekciu reflexov nad vnútornou pamäťou, ktorá má určitú špeciálnu povahu** (potenciálne ohraničenú časovú platnosť a priority).

Poďme sa teraz pozrieť ako to bude vyzerať na zvolenom príklade. V modeli budeme mať niekoľko globálnych premenný reprezentujúcich vnímanie kutavky:

nestDetected	Kutavka sa nachádza pri svojej komôrke
victimDetected	Kutavka vidí sedlovku
paralyzedVictimDetected	Kutavka vidí sedlovku, ktorá sa nehýbe
insideNest	Kutavka vidí, že je vo vnúti komôrky
nestClosed	Kutavka vidí uzavretú komôrku

ktoré dopĺňa premenná reprezentujúca vnímanie vlastného tela:

inseminated	Kutavka je oplodnená

Ďalej budeme mať premenné reprezentujúce vnútornú pamäť kutavky:

nestReady	komôrka je vykopaná
checkingPerformed	kontrola komôrky je vykonaná
checking	kontrola komôrky prebieha
victimInNest	sedlovka je v komôrke

a na niekoľko základných správaní (ktoré by sme si ale v detailoch tiež vedeli predstaviť ako zložené z atomickejších úkonov realizovaných podobnými agentovými štruktúrami – na to by sme však museli uvažovať jemné detaily tela kutavky, čo by v tejto chvíli bolo zbytočne zložité):

digNest()	kop komôrku
lookForVictim()	hľadaj sedlovku
paralyzeVictim()	paralyzuj sedlovku
carryToNest()	zanes paralyzovanú sedlovku ku komôrke
checkNest()	kontroluj komôrku
pullIntoNest()	vtiahni paralyzovanú sedlovku do komôrky
closeNest()	uzavri vchod do komôrky

Pod ich „zavolaním“ si musíme predstaviť vypustenie výstupu z daného správania na aktuátory kutavky na relatívne krátky čas, presnejšie pokým nejaký iný agent „nezavolá“ správanie iné. Ak teda v kóde „zavoláme“ digNest() myslíme tým len malý krôčik v kopaní hniezda a „volanie“ agenta sa vôbec nezablokuje (tradične by sme predpokladali,

že ho zablokuje, až kým sa nevykope celé hniezdo). Modelované správanie kutavky potom možno vysvetliť súčinnosťou siedmich agentov:

Agent1

```
for (each 1 second)
  if (inseminated && !nestReady) {
    if (insideNest) nestReady = true for 1 day;
    else digNest();
  }
```

Agent2

```
for (each 1 second)
  if (nestReady && !victimDectected)
    lookForVictim();
```

Agent3

```
for (each 1 second)
  if (victimDectected && !paralyzedVictimDectected)
    paralyzeVictim();
```

Agent4

```
for (each 1 second)
  if (paralyzedVictimDectected && !nestDectected)
    carryToNest();
```

Agent5

```
for (each 1 second)
  if (!checkingPerformed && ((paralyzedVictimDectected
     && nestDectected) || checking) {
    checking = true for 7 seconds;
    checkNest();
    if (insideNest)
      checkingPerformed = true for 40 seconds;
  }
```

Agent6

```
for (each 1 second)
  if (checkingPerformed && paralyzedVictimDetected
      && !victimInNest)
    if (insideNest) victimInNest = true for 1 hour;
    else pullIntoNest();
```

Agent7

```
for (each 1 second)
  if (victimInNest & !nestClosed) closeNest();
```

V tomto modeli už pravda nenahliadneme tak ľahko, čo sa vlastne za akých podmienok robí a ak sa nechceme pomýliť, treba prejsť ku počítačovej simulácii. V princípe sa ale za normálnych okolností agenty aktivizujú v uvedenom poradí, t.j. istý čas je aktívny agent1, potom agent2, atď. Avšak spôsob ako jeden agent aktivizuje druhého je rôzny od prípadu k prípadu.

Niekedy je aktivácia sprostredkovaná následkom činnosti predchádzajúceho agenta, napríklad paralyzovanie obete sa spúšťa následkom toho, že ju nájdeme. V tomto prípade teda nie je potrebná vnútorná pamäť.

V iných prípadoch treba aktiváciu realizovať cez vnútornú pamäť, ale pokiaľ ide o dlhodobo platný údaj, ako napríklad, že komôrka je vykopaná, dá sa to ešte vyjadriť dosť jednoducho.

Najzložitejšie vyjadrenie vyžaduje práve tá inkriminovaná kontrola komôrky, ktorá sa dá oprieť výlučne o vnútornú pamäť, lebo prostredie sa jej vykonaním nemení a navyše ide o krátkodobo platné zistenie. Musí sa tam nejako zabezpečiť, aby kutavka po istom čase kontrolu znovu vykonala a zabrániť, aby tak neurobila ihneď po skončení predošlej kontroly. Navyše túto fázu štartujú podmienky, ktoré už tesne po jej zahájení prestanú platiť (ulovenú sedlovku necháme pri kontrole vonku). To sa najjednoduchšie dosiahne použitím časovej platnosti premenných. Akákoľvek eliminácia časových platností tu vyžaduje implementáciu zabúdania, a povedie teda k zložitejšiemu kódu v agentoch. Tupé správanie kutavky sa tu teda dá vysvetliť na základe toho, že kým kutavka sedlovku opäť nájde, zabudne, že komôrku skontrolovala. Príroda si vybrala túto možnosť zrejme preto, že jej kód je jednoduchší, hoci logicky nie je stopercentne správny. Navyše treba vziať v úvahu, že sa uvažované zlyhanie prejavuje za podmienok, ktoré sa v prírode nevyskytujú, a teda nemôže nijako ovplyvniť evolučný výber. Zbytočnou kontrolou v nastrojených podmienkach sa tu teda asi platí za jednoduchosť riadiacich štruktúr postačujúcich v prirodzených podmienkach.

Čo sa týka vysvetlenia, prečo po čase – hoci dlhom čase - kontrolu predsa len kutavka zanechá, v rámci tohto modelu možno odpovedať tým, že opätovné nájdenie sedlovky netrvá vždy rovnako dlho a opakovaním sa môže zdokonaľovať. Tým sa môže podariť kutavke vrátiť sa ku komôrke v čase, keď ešte časová platnosť kontroly nevypršala. (Navyše, keď zvážime biologické možnosti implementácie časovej platnosti – časté zapisovanie takého údaju by zrejme viedlo k posilneniu schopnosti zapamätať si ho na dlhší čas a tým pádom by sa po väčšom počte opakovaní platnosť kontroly dostatočne predĺžila.)

Výhodou tohto modelu je, že okrem zhodovania sa s normálnym správaním i správaním počas uvažovaného experimentu dáva dobré predpovede pre iné okrajové situácie. Predovšetkým zahŕňa v sebe návrat k vhodným predošlým fázam v prípade výskytu výnimky. Napríklad, keď počas prinášania kutavke sedlovka vypadne, model automaticky prejde do fázy hľadania. Ak pri opätovnom prinášaní nájde inú – ešte neulovenú – sedlovku, paralyzuje ju. Ak však nájde tú istú, paralyzovanie vynechá. V kritickom prechode od kontroly k vtiahnutiu sa na základe straty sedlovky vráti do fázy hľadania.

Vieme však urobiť aj zaujímavé predpovede, u ktorých nevieme odhadnúť ako by dopadlo ich otestovanie v prírode? Hneď niekoľko:

1. čo sa stane, keď pri kontrole ochrnutú sedlovku odstránime a dáme miesto nej inú neochrnutú, ale zato napríklad uviazanú na nitku aby neušla? Z modelu vychádza, že by ju mala paralyzovať a vtiahnuť.
2. aký vplyv bude mať vzdialenosť odtiahnutia sedlovky na počet opakovaní? Tu nám vychádza, že ak bude vzdialenosť dostatočne malá, žiadne opakovanie sa nedostaví. Inak vo všeobecnosti, čím menšia vzdialenosť, tým menej opakovaní.
3. dá sa zväčšiť počet opakovaní z tých 30-40? Podľa modelu by sa to malo podariť tak, že budeme stále vhodne zväčšovať vzdialenosť, o ktorú sedlovku odťahujeme.
4. čo sa stane keď komôrku zakopeme? Tu sa zo značnej časovej platnosti údaju o vykopaní komôrky dá usúdiť, že kutavka sa dosť dlho bude správať načisto bezradne – dlho bude hľadať neexistujúcu komôrku. Nakoniec si však predsa len vykope komôrku novú.
5. Bude sa pri opakovaniach prinesenia sedlovky čas tohto prinesenia skracovať (tréningom) alebo predlžovať (únavou)? Pre náš model by bolo priaznivé keby sa rekordný čas stále zmenšoval (kolísať pritom môže koľko chce).

Keby nejaký etológ zistil – či už z existujúcich prác alebo vykonaním experimentu, nakoľko sú tieto predpovede úspešné, dozvedeli by sme sa nielen či sme spravili dobrý model, ale by sme aj posilnili či oslabili našu vieru, že správne hádame s akými štruktúrami príroda pracuje.

5 Záver

V tomto príspevku sme sa snažili čitateľa presvedčiť, že pri modelovaní živých systémov v počítači by sme sa nemali zameriavať na ich vonkajšie prejavy, ale na ich vnútorné štruktúry, ktoré sú za ne zodpovedné. Žiadali sme pritom, aby sme modelovaním skúmali povahu týchto štruktúr podobným spôsobom ako sa vo fyzike skúma štruktúra hmoty, predovšetkým aby tieto modely poskytovali dostatočne zaujímavé predpovede, ktoré sa potenciálne dajú otestovať.

Na príklade z etológie hmyzu sme demonštrovali vlastnú predstavu o povahe týchto štruktúr, ktorá je síce len jednou z mnohých, ale ktorá sa vyznačuje zaujímavým potenciálom. Otvorili sme také základné otázky z počítačovej simulácie ako „čo je premenná?“ a „ako jeden kus kódu volá druhý?“. Ukázali sme aký význam má uvažovať u premennej obmedzenú časovú platnosť a prioritu a aký význam má modularita založená na agentoch. V podstate, ak je vôbec slušné povedať, že príroda programuje svoje výtvory, tak my k tomu dodávame, že používa trochu iné premenné a inak organizovaný kód, než si bežne pod programovaním predstavujeme. S tým asi bude súhlasiť takmer každý, my ale máme aj určitú konkrétnu predstavu aký charakter tieto premenné a kód majú.

Ako vedľajší efekt sme sa snažili ukázať, že niektoré zaujímavé vlastnosti získavajú živé systémy prostoduchejším spôsobom, než by sme predpokladali. Doslova ich majú vďaka svojej nedokonalosti. Jednoznačným dôsledkom tohto faktu je, že pokiaľ chceme zostrojiť stroje s podobnými vlastnosťami, mali by sme túto nedokonalosť použiť ako konštrukčný princíp. Toto zároveň poukazuje na rozdiely medzi strojmi a živými tvormi. Vo svojom správaní sa zrejme nepodobajú práve preto, že pri konštrukcii našich strojov sa snažíme o ich dokonalosť.

Literatúra

[1] Brooks, R.: *Cambrian Intelligence.* Cambridge, MA: The MIT Press 1999.
[2] Brooks, R.: *Robot – The Future of Flesh and Machines.* London: Penguin Books 2002.
[3] Gál, E.: *Intuitívna psychológia a kognitívne vedy.* In: Kognitívne vedy III (Kvasnička V., Pospíchal J., eds.), CHTF, STU, Bratislava, 2000.
[4] Hass, H.: *The Human Animal: The Mystery of Man's Behavior.* G. P. Putnam's Sons 1970, pp. 27
[5] Kelemen, J.: *Strojovia a agenty.* Bratislava: Archa 1994.
[6] Kelemen, J.: The Agent Paradigm. *Computing and Informatics* **22** (2003) 513-519.
[7] Lúčny, A.: *Hľadanie kvalitatívneho rozdielu.* In: Kognice a umělý život (Kelemen, J. – Kvasnička, V. – Pospíchal, J., eds.), FPF SU, Smolenice, 2001, 167-176.
[8] Lúčny, A.: Building Complex Systems with Agent-Space Architecture. *Computing and Informatics* **23** (2004) 1001-1036.
[9] Minsky, M.: *The Society of Mind.* New York: Simon&Schuster 1986.
[10] Minsky, M.: *Emotion machine.* http://web.media.mit.edu/~minsky/, 2004.

Jazyk ako graf

Mária Markošová, Peter Náther[1]

Abstrakt. Siete malého sveta majú takú štruktúru, ktorá integruje vlastnosti usporiadaných a náhodných grafov. Reálne siete vznikajúce prirodzeným rastom, ako sú napríklad komunikačné siete, sociálne siete, neurónové siete, možno modelovať pomocou sietí malého sveta. V tejto práci ukazujeme, že aj lexikón jazyka má charakter siete malého sveta a to na príklade anglického aj slovenského jazyka. Chceme tiež ukázať, aký vplyv má spôsob rastu siete na jej výslednú štruktúru.

1 Čo sú siete malého sveta?

Pojem *sieť malého sveta* bol vytvorený psychológmi [1]. Autori sa zaoberali štúdiom štruktúry sociálnych sietí a na základe svojich pozorovaní odvodili *zákon šesťstupňovej separácie,* ktorý možno sformulovať napríklad takto: *Medzi dvoma náhodne vybranými uzlami sociálnej siete je v priemere šesť hrán.*

Čo to vlastne znamená? Autori článku [1] urobili v USA experiment, v ktorom skúmali, ako sa ľudia cez osobné kontakty dokážu dostať jeden k druhému. Zistili, že ak v Spojených štátoch amerických náhodne volíme dvojice obyvateľov a prikážeme im, aby sa medzi sebou skontaktovali cez reťaz svojich známych, táto reťaz je v priemere veľmi krátka. Jeden človek sa k druhému dostane na šesť krokov.

Zákon šesťstupňovej separácie je len inou formuláciou tzv. Kevin Bacon efektu, pozorovaného v sieti hercov [2]. Kevin Bacon je menej známy americký herec. Priraďme mu baconovské číslo $C_b = 0$. Predstavme si, že herec, ktorý hral v akomkoľvek filme s Kevinom Baconom má $C_b = 1$. Ak nejaký herec hral vo filme s niekým, kto hral vo filme s Kevinom Baconom, má $C_b = 2$; ak niekto hral vo filme s hercom, ktorý hral vo filme s niekým, kto hral s Kevinom Baconom má baconovské číslo $C_b = 3$. A tak podobne. Ukazuje sa, že rozhodujúca časť hercov, ktorí kedy hrali v nejakom americkom filme, nemá baconovské číslo väčšie ako 4. Najvyššie doteraz zistené baconovské číslo bolo osem. Je tomu tak preto, že profesionálna sieť hercov má charakter siete malého sveta, a teda aj malú separáciu uzlov (hercov) siete.

Pod separáciou uzlov l rozumieme priemernú najkratšiu vzdialenosť medzi dvoma náhodne zvolenými uzlami a to meranú počtom hrán ležiacich medzi nimi. Je to veličina, ktorú v sieti môžeme zmerať. V usporiadaných mriežkových grafoch malej dimenzie d táto veličina pomerne rýchlo rastie s počtom uzlov grafu N [2,3]:

$$l \propto N^{1/d} \tag{1}$$

V prípade náhodných grafov platí [2,3]:

$$l \propto log(N) \tag{2}$$

teda separácia uzlov rastie s počtom uzlov omnoho pomalšie ako pre mriežkové grafy.

[1] Fakulta matematiky, fyziky a informatiky, Katedra aplikovanej informatiky, Mlynská dolina, Bratislava, E-mail: markosova@ii.fmph.uniba.sk, nather@ii.fmph.uniba.sk

Ďalšou merateľnou veličinou v sieťach je koeficient klasterizácie C [2,3]. Každý uzol siete je pospájaný so svojimi susedmi pomocou hrán. Nech uzol i má k_i bezprostredných susedov, ktorí sú s ním spojení hranami. Táto veličina máva v literatúre rôzne mená, nazýva sa konektivitou uzla i, alebo stupňom uzla i. Ale aj susedia uzla i môžu byť poprepájaní medzi sebou, môžu tvoriť klaster. Matematicky je teda koeficient klasterizácie i-teho uzla daný ako pravdepodobnosť toho, že dvaja najbližší susedia uzla i sú aj vzájomne najbližšími susedmi. Teda

$$C_i = \frac{2E_i}{k_i(k_i - 1)} \tag{3}$$

je klasterizačný koeficient uzla i. E_i je počet reálne existujúcich hrán medzi susedmi uzla i, a $\frac{k_i(k_i-1)}{2}$ je počet možných hrán medzi k_i bodmi, keby boli poprepájané spôsobom každý s každým. Klasterizačný koeficient celej mriežky je priemerom klasterizačných koeficientov jednotlivých uzlov, teda

$$C = \frac{1}{N}\sum_{i=1}^{N} C_i \tag{4}$$

kde N je počet uzlov mriežky.

Siete malého sveta sú pozoruhodné tým, že majú malú separáciu uzlov, porovnateľnú s náhodnými grafmi, ale často krát veľkú klasterizáciu, porovnateľnú s usporiadanými mriežkovými grafmi s klasterizáciou. Náhodné grafy majú klasterizačný koeficient blízky k nule.

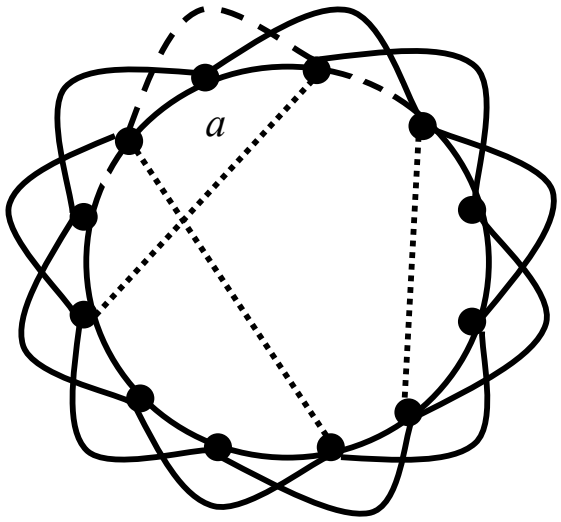
Dôvodom týchto vlastností je štruktúra sietí malého sveta. V stručnosti možno povedať, že sa vlastne často krát jedná o usporiadané mriežkové grafy s niekoľkými náhodnými prepojeniami uzlov. Typickým príkladom siete malého sveta je Newmanov-Wattsov model (Obr.1) [3]. Je to jednorozmerná usporiadaná mriežka, v ktorej hrany existujú nielen medzi najbližšími, ale aj medzi druhými najbližšími susedmi. To zavádza do mriežky klasterizáciu. Potom na tejto mriežke definujeme takýto proces:

1. Pohybujeme sa v smere hodinových ručičiek po hranách spájajúcich najbližších susedov.
2. Každú hranu s pravdepodobnosťou p presmerujeme k náhodne zvolenému uzlu v mriežke. V mriežke vznikne krátke spojenie.
3. Keď ošetríme hrany spájajúce najbližších susedov, opakujeme od bodu 1 pre druhých najbližších susedov.

Teoretické výpočty ukazujú, že ak je $d = 1$, počet susedov uzla je párny a väčší alebo rovný dvom, potom

$$C = \frac{3}{4}\frac{(k-2)}{(k-3)} \tag{5}$$

kde k je počet susedov uzla na pravidelnej mriežke. Ak je k veľké, potom $C \to \frac{3}{4}$ [2].



Obrázok 1. Newmanov – Wattsov model s klasterizáciou. Bod *a* má štyroch susedov, ktorí sú medzi sebou poprepájaní prerušovanými hranami. Náhodné krátke spojenia sú vyznačené bodkovanou čiarou.

2 Lexikón jazyka ako sieť malého sveta

Lexikón ľudského jazyka je zložený z niekoľkých desiatok tisíc slov. Napriek tomu sa ľudský mozog v tejto rozsiahlej databáze veľmi dobre orientuje. To okamžite vyvoláva otázku, ako je slovník implementovaný v mozgu. Jedna z teórií tvrdí, že lexikón jazyka tvorí sieť, ktorá má charakter siete malého sveta [4 - 6].

Predstavme si, že každé slovo je uzlom slovnej siete (word web). Nech interakcie medzi slovami tvoria hrany. Je to zaujímavá predstava, v ktorej však vznikajú aj problémy. Napríklad, čo je to vlastne slovo? Ako ho chápať? V rôznych jazykoch je to rozdielne. Ak máme namysli angličtinu, vec je vcelku dosť jednoduchá, pretože angličtina prakticky nepozná rozličné tvary toho istého slova. Horšie je definovať slovo napr. v slovenskom jazyku, kde pomocou predpôn a prípon možno utvoriť rôzne, často krát významovo dosť odlišné slová. Napriek tomu sú však odvodeninou základného tvaru. Napríklad také slovo **rod.** Odvodeniny sú od**rod**a, **rod**ný, zá**rod**ok, **rod**ič, **rod**isko... Brať všetky tieto slová ako rôzne, alebo ako jedno slovo, hoci trebárs odroda a rodič označujú veci, ktoré nemajú významovo až tak veľa spoločného.

Ďalší problém je definovať interakciu medzi slovami, teda hranu grafu. Ako chápať interakciu? Syntakticky, alebo sémanticky? V rozličných štúdiách slovných sietí, ktoré boli doteraz spracované, je interakcia definovaná oboma spôsobmi. Motter a kol. [5] skúmali tzv. konceptuálnu sieť, v ktorej sa pod interakciou dvoch slov rozumie to, že jedno slovo sa nachádza v kontexte iného slova. V praxi sa to robí napríklad tak, že sa slová považujú za prepojené hranou vtedy, ak jedno slovo je v klastri slov popisujúcich druhé slovo v nejakom synonymickom, alebo výkladovom slovníku. Motter a jeho spoluautori použili na sieti dostupný Thesaurus dictionary. Cancho a Solé [4] a tiež Dorogovtsev a Mendes [6] volili iný pohľad, ktorý je bližšie k syntaktickej stavbe jazyka. Ak sú slová susedmi daného slova vo vete, považujú sa s ním za spojené hranou. Na

základe týchto dvoch pohľadov možno vytvoriť dva typy slovných grafov, nazvime ich *konceptuálnym* a *polohovým* grafom.

2.1 Konceptuálna sieť

Prvé modely slovných sietí vychádzajú z angličtiny. Je to jazyk morfologicky jednoduchý a dá sa dobre spracovať na počítači. Konceptuálny graf angličtiny ako prví spracovali Motter a kol. [5]. Za základ si zobrali Thesaurus Dictionary. V tomto slovníku je 30 000 kľúčových slov spracovaných tak, že po zadaní požiadavky na vyhľadanie slova v základnej databáze dostaneme aj slová s ním pojmovo spojené. Tie sa potom považovali za slová spojené hranou s pôvodným slovom, ale len vtedy ak sa nachádzali medzi 30 000 kľúčovými slovami. Tie, ktoré nespĺňali túto podmienku, sa nebrali do úvahy. Takto možno vytvoriť konceptuálny slovný graf. Numerická analýza tohto grafu ukázala, že každé slovo má v priemere 60 spojení s inými slovami.

Počítačové spracovanie anglického lexikónu ukázalo, že konceptuálna sieť jazyka je veľmi klasterizovaná a riedka a má vlastnosti siete malého sveta, teda dosť veľký klasterizačný koeficient $C = 0.53$, doprevádzaný malou separáciou uzlov, približne rovnou trom. Teda, ak v jazykovom grafe náhodne vyberieme dve slová, sú v priemere spojené cez tri ďalšie hrany grafu. V tak veľkej sieti to naozaj nie je veľa. V tabuľke 1 sa nachádza porovnanie parametrov konceptuálneho grafu lexikónu anglického jazyka s parametrami náhodného grafu s rovnakým počtom uzlov N i rovnakým priemerným počtom hrán $\overline{k}$ vychádzajúcich z jedného uzla. Dôvodom krátkej vzdialenosti medzi dvoma slovami sú tie slová, ktoré patria do dvoch alebo viacerých rôznych kontextov, a vytvárajú tak ďalekodosahové interakcie medzi kontextovými klastrami. Patrí k nim napríklad slovo *nature*, ktoré v angličtine znamená príroda, ale napríklad aj povaha.

Tabuľka 1. Porovnanie parametrov konceptuálnej siete jazyka s parametrami náhodného grafu s rovnakým počtom uzlov N a rovnakým priemerným počtom hrán $\overline{k}$. Podľa [5].

	N	$\overline{k}$	C	l
Kontext. sieť	30 244	59.9	0.53	3.16
Náhodný graf	30 244	59.9	0.002	2.5

2.1 Polohová sieť

Ak sa na lexikón jazyka pozeráme zo syntaktického hľadiska, môžeme vytvoriť dva typy grafov. Ak berieme do úvahy len také spojenia slov, ktoré sa vyskytujú častejšie ako pri náhodnom radení slov do vety, teda ak uplatníme podmienku

$$p_{ij} \succ p_i p_j \tag{6}$$

dostaneme RWN (restricted word net). V (6) p_i, p_j sú pravdepodobnosti výskytov slov *i, j* v skúmaných textoch a p_{ij} pravdepodobnosť výskytu *i*-teho a *j*-teho slova vedľa seba. V prípade, že podmienku (6) nezahrnieme, máme UWN (unrestricted word web) [4]. Numerická analýza textov v anglickom národnom korpuse ukázala, že aj takáto sieť má vlastnosti siete malého sveta, teda veľký koeficient klasterizácie a malú separáciu uzlov. Získané numerické dáta, týkajúce sa anglického jazyka, sú zhrnuté v tabuľke 2.

Tabuľka 2. Porovnanie vlastností UWN a RWN polohových slovných sietí s vlastnosťami náhodných grafov s rovnakým počtom uzlov $N_{UWN} = 478\ 773$, $N_{RWN} = 460\ 902$ a priemerným počtom hrán /uzol $\overline{k}_{UWN} = 74{,}2$, $k_{RWN} = 70{,}13$. C je koeficient klasterizácie, l je separácia uzlov, C_{random}, l_{random} sú tie isté veličiny v náhodnom grafe. Podľa [4]

Typ grafu	*C*	C_{random}	*l*	l_{random}
UWN	0.687	1.55×10^{-4}	2.63	3.03
RWN	0.437	1.55×10^{-4}	2.67	3.06

Ako sme už spomenuli vyššie, slovenčina je jazyk podstatne náročnejší na počítačové spracovanie, pretože má bohatú morfologickú štruktúru [9]. V angličtine význam slova vo vete dosť závisí od jeho polohy vo vete a od slovosledu. V slovenčine je slovosled voľnejší a význam slova je viacej určený jeho tvarom. Všimnime si napríklad vety: *„He is a world champion in chess“. („On je svetový majster v šachu“), „This is a beautifull world“ („Tento svet je nádherný“).* V angličtine pojmy *svetový* a *svet* sú symbolizované jedným slovom *world*, rozlíšenie prídavného a podstatného mena je dané polohou vo vete. Na rozdiel od toho v slovenčine prídavné meno tvoríme príponou. Aj veta *„On je naprosto svetový“* má preto zmysel.

Aké tvary slov máme teda v slovenčine považovať za uzly grafu? V práci [9] bol zvolený taký prístup, že gramatické kategórie odvodené z jednej koreňovej morfémy boli považované za jedno slovo. Za základ analyzovaného slovenského textu slúžili rôzne články stiahnuté z Internetu, ktoré autor používal ako testovací materiál na vývoj programov pre automatické spracovanie slovenského jazyka. Výsledky spracovania ukazujú, že aj polohová slovná sieť vytvorená na báze slovenského jazyka má charakter siete malého sveta [9]. V tabuľke 3 je porovnanie vlastností polohových slovných sietí G_1 a G_2. V grafe G_1 slovo (uzol) je spojené väzbou len s najbližšími susedmi vo vete, v grafe G_2 aj s druhými najbližšími susedmi. Pre porovnanie sú uvedené aj vlastnosti náhodného grafu s rovnakými parametrami (N, $\overline{k}$).

Tabuľka 3. Porovnanie parametrov polohovej siete slovenského jazyka s parametrami náhodného grafu s rovnakým počtom uzlov $N=59\ 542$ a rovnakým priemerným počtom hrán $\overline{k}$. Význam parametrov v tabuľke 3 je taký istý ako v tabuľke 2. Podľa [9].

Typ grafu	*C*	C_{random}	*l*	$\overline{k}$
G_2	0.607	8.9×10^{-4}	2.62	53.01
G_1	0.369	3.36×10^{-4}	2.87	29.96

Porovnanie tabuliek 2 a 3 ukazuje, že vlastnosti grafov zostavených na báze anglického a slovenského jazyka sú veľmi podobné. Obe siete kombinujú vysokú hodnotu klasterizačného koeficientu s malou separáciou uzlov, a sú teda sieťami malého sveta.

3 Dynamika sietí

Klasterizačný koeficient C (3,4) a separácia uzlov l sú veličiny, pomocou ktorých môžeme charakterizovať siete, ktoré sa s časom nemenia, nerastú a ani sa počet ich uzlov nezmenšuje. Ak pomocou týchto veličín charakterizujeme siete, v ktorých sa počet uzlov mení, tak to môžeme urobiť len za predpokladu, že sme zastavili čas. Avšak všetky siete, vznikajúce prirodzenými procesmi, napríklad komunikačné siete, sociálne siete, neurónové siete a podobne, sa s časom menia, majú istú dynamiku. Aj do lexikónu slová neustále pribúdajú; a samozrejme, niektoré slová sa s časom prestávajú používať. Je preto veľmi dôležité rozumieť aj dynamike sietí. Z hľadiska jazyka sú dôležité najmä rastúce siete, pretože počet používaných slov sa s časom zväčšuje.

Rastúce siete študovali Barabási a Albert [7]. Navrhli dva modely a skúmali, ako spôsob pripájania sa nových uzlov ovplyvňuje celkovú štruktúru siete v čase t. Prvý model, nazývaný *model s náhodným pripájaním uzlov,* je takýto:

1. Každú časovú jednotku prichádza do siete nový uzol.
2. Vytvorí hranu s už existujúcim uzlom a to tak, že pravdepodobnosť vytvorenia tejto hrany je rovnaká pre všetky staré uzly v sieti.

Druhý model rastúcej siete, ktorý sa volá *model s preferenčným pripájaním uzlov,* je takýto:

1. Každú časovú jednotku prichádza do siete nový uzol.
2. Nový uzol vytvorí hranu s jedným zo starých uzlov a to tak, že pravdepodobnosť vytvorenia tejto hrany je úmerná počtu hrán vychádzajúcich zo starého uzla.

Ako ukázali autori práce [7], spôsob pripojenia nových uzlov ovplyvňuje štruktúru siete. Je to možné vidieť napríklad z tvaru funkcie, ktorá sa volá distribúcia konektivity. Na osi x je konektivita k (stupeň uzla, teda počet hrán vychádzajúcich z uzla) a na osi y počet uzlov (poprípade znormovaný), ktoré majú stupeň k. Túto funkciu možno v sieti určiť meraním. Štúdia [7] ukazuje, že na jej základe možno usúdiť aká bola história siete. Pretože ak sa uzly pripájali náhodným spôsobom, distribučná funkcia má tvar

$$P(k) \propto e^{-\beta k} \tag{7}$$

Ak sa uzly k sieti pripájali preferenčným spôsobom, potom funkcia $P(k)$ vyzerá takto:

$$P(k) \propto k^{-\gamma}, \tag{8}$$

kde možno ukázať, ako uvidíme ďalej, že γ je rovná trom. Preferenčný spôsob pripájania je zaujímavejší, pretože takýmto procesom, ako vidíme z tvaru funkcie (8), vynikajú tzv. *bezškálové siete*. Tie sú významné preto, že ich nie je ľahké vymazaním malého počtu náhodne vybraných uzlov rozdeliť na časti. Sú teda odolné voči náhodným poruchám.

Dorogovtsev a Mendes [10] študovali Barabási - Albert model s preferenčným pripájaním uzlov. Vytvorili spojitý model, diferenciálnu rovnicu popisujúcu procesy v sieti, ktorej riešením možno získať číselnú hodnotu exponentu γ a aj tvar distribučnej funkcie (8). Nech $k(s,t)$ je stupeň (konektivita) uzla, ktorý prišiel do sieti v čase s a pozorujeme ho v čase t. Potom sa konektivita tohto uzla v čase vyvíja takto

$$\frac{dk(s,t)}{dt} = \frac{k(s,t)}{\int_0^t du \quad k(u,t)}. \tag{9}$$

Integrál v menovateli je normovacia konštanta. Riešením tejto rovnice je $k(s,t) = \left(\frac{t}{s}\right)^{1/2}$.

Použitím vzťahu

$$P(k,t) = -\left[t\frac{\partial k(s,t)}{\partial s}\right]^{-1}_{s=s(k,t)} \tag{10}$$

vypočítame $P(k) = 2k^{-3}$, čo je vzťah (8) s $\gamma = 3$.

3.1 Dynamika jazykového grafu

Pozrime sa najprv na konceptuálny slovný graf. Podľa [5] sa nový uzol pripája do konceptuálnej siete aj preferenčným aj náhodným spôsobom. Numerické výsledky naznačujú, že ak sú uzly málo prepojené, nový uzol sa k nim pripojí skôr náhodným spôsobom. Avšak uzly, ktorým sa podarí získať veľa spojení, majú šancu zachytiť ďalšie uzly s istou preferenciou. Toto je reflektované v tvare distribučnej funkcie, ktorá má pre malé hodnoty k exponenciálny charakter, ale pre veľké hodnoty k prechádza do mocninného tvaru (8). Motter a spol. [5] usudzujú, že tento tvar distribučnej funkcie vzniká vďaka dvojstupňovému procesu pripájania uzlov. Sieť má *small world* topológiu, teda sa skladá z klastrov uzlov, majúcich mnoho lokálnych spojení a málo globálnych, ďalekodosahových spojení. Keď do takejto siete príde nový uzol, pripojí sa s rovnakou pravdepodobnosťou ku ktorémukoľvek staršiemu uzlu. Ale ak už raz nový uzol vytvoril hranu napríklad s uzlom *j*, potom má tendenciu preferenčne si vytvárať spojenia s uzlami, ktoré už sú spojené s uzlom *j*.

Ak interakciu slov chápeme syntakticky (polohová slovná sieť), potom dominuje preferenčné pripájanie nových uzlov do siete, ktoré sa odráža v mocninnom tvare distribúcie konektivity (8). Situácia, ako ukazuje Cancho a Solé na príklade anglického jazyka [4], je však o niečo zložitejšia. V distribúcii (8) možno pozorovať dva režimy. V prvom, kde $k \le k_{cross}$, teda relatívne malé, má nameraný exponent γ hodnotu *1.5,* v druhom, pre $k \succ k_{cross}$, $\gamma = 2.7$. Táto druhá hodnota exponentu γ je blízka teoreticky vypočítanej hodnote *3*, teda exponentu získaného riešením Barabási – Albert modelu [7], (9,10). Barabási – Albert model popisuje preferenčné pripájanie uzlov do dynamicky ustálenej siete rastúcej s časom. Podľa [4] dva škálovacie režimy možno vysvetliť tým, že dobre prepojené slová v jazykovom grafe sú slová z jadra jazyka, slová najčastejšie používané, zrozumiteľné všetkým dospelým ľuďom hovoriacich daným jazykom, bez ohľadu na vek, vzdelanie a podobne. V jadre jazyka je situácia taká, ako v Barabási – Albert modeli. Zatiaľ čo slová z periférie jazyka, ktoré sa používajú obmedzene, a teda nie sú v slovníku väčšiny, majú inú dynamiku.

Dorogovtsev a Mendes [6] vytvorili kontinuálny model, diferenciálnu rovnicu, ktorej riešením sú vyššie popísané dva režimy. Použili klasický Barabási – Albert model s preferenčným pripájaním (9), ktorý modifikovali takto:

1. Každú časovú jednotku do siete prichádza jeden uzol.

2. Tento uzol sa preferenčným spôsobom pripojí k existujúcim uzlom v sieti a to tak, že pravdepodobnosť vytvorenia linky medzi novým a starým uzlom je úmerná počtu hrán vychádzajúcich zo starého uzla.
3. Zároveň vznikne preferenčným spôsobom ct nových hrán medzi starými uzlami, pričom $1 \prec\prec ct \prec\prec t$.

Model vyzerá takto:

$$\frac{\partial k(s,t)}{\partial t} = (1+ct)\frac{k(s,t)}{\int_0^t du\, k(u,t)} \tag{11}$$

Riešením rovnice (11) je

$$k(s,t) = \left(\frac{ct}{cs}\right)^{1/2}\left(\frac{2+ct}{2+cs}\right)^{3/2} \tag{12}$$

Dá sa ukázať [6,10], že $k(s,t) \propto s^{-\beta}$ a tiež, že

$$\gamma = 1 + \frac{1}{\beta} \tag{13}$$

Ak vezmeme do úvahy, že staršie uzly sú lepšie prepojené a sú to práve tie, ktoré prišli do siete v čase $s \prec\prec t$ a novšie a slabšie prepojené uzly sú tie, ktoré došli do siete v čase $s \approx t$ a zároveň vezmeme do úvahy vzťahy (12), (13), zistíme, že

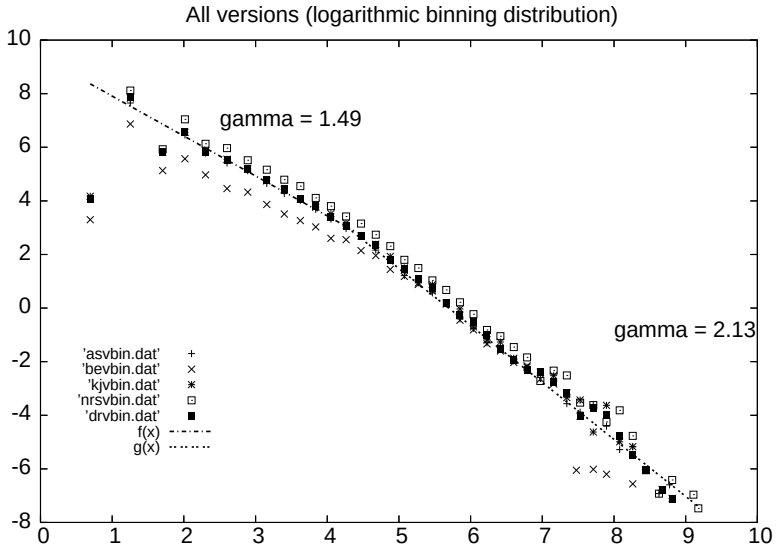
$$\gamma = \frac{1}{1+\frac{1}{2}} = 3, \text{ alebo } \gamma = \frac{1}{\frac{1}{2}+\frac{3}{2}} = 1/2 \tag{14}$$

Riešením Dorogovtsevovho a Mendesovho modelu sú teda dva škálovacie režimy s dvoma rôznymi exponentami $\gamma = 3$ pre dobre prepojené uzly a $\gamma = 1/2$ pre uzly, ktoré sú prepojené horšie. Prechodový bod medzi dvoma režimami je [6] $k_{cross} \approx \sqrt{ct}(2+ct)^{3/2}$.

3.2 Experimentálne výsledky

Aby sme overili hypotézu Cancha a Solého [4], totiž, že lepšie prepojené slová v polohovej slovnej sieti sú slová z jadra jazyka a ich funkcia *P(k)* sa správa inak ako pre horšie prepojené slová, vytvorili sme slovnú sieť pre viacero anglických prekladov textu, ktorý, ako sme sa domnievali, je zložený hlavne zo slov z jadra jazyka. Týmto textom je Biblia. Na sieti [11] sme našli rôzne verzie anglického prekladu Biblie, z ktorých sme vytvorili polohové slovné grafy.

Na obrázku 2 je distribúcia konektivity pre všetky verzie anglických prekladov Biblie, ktorých charakteristiky sú v tabuľke 4.



Obrázok 2. Nenormovaná distribučná funkcia konektivít pre slovnú sieť zostavenú pre všetky preklady Biblie, ktorých charakteristiky sú v tabuľke 4. Na osi x je *ln(N)* a na osi y *ln(k),* kde *N* je počet slov a *k* je konektivita.

Tabuľka 4. Charakteristiky anglických prekladov Biblie. DRV – Douhay Rheims verzia, KJV – King James verzia, ASV – American Standard verzia, BEV – Basic English verzia, NRSV – New Revisited Standard verzia. EME znamená Early Modern English, SE – Simplified English a ME – Modern English.

verzia	rok vydania	počet rôznych slov	jazyk (verzia angličtiny)
DRV	*1582*	*14513*	EME
KJV	*1611*	*14518*	EME
ASV	*1901*	*12749*	EME
BEV	*1941*	*6237*	SE
NRSV	*1989*	*12749*	ME

Z obrázku 2 vidíme, že tvar funkcie napovedá, že slovná sieť rástla preferenčným pripájaním uzlov. Napriek nášmu predpokladu, že text sa skladá zo slov, ktoré sú prevažne slovami z jadra jazyka, a teda sme očakávali, že nameraný exponent γ bude blízky trom, dva režimy, s dvoma rôznymi exponentami ostali zachované. Je to prekvapujúci výsledok, ktorý si žiada širšiu diskusiu.

4 Diskusia

Čo vyjadruje štruktúra slovnej siete? Odpoveď na túto otázku zatiaľ nie je úplná a vyžaduje si ďalší výskum. Predovšetkým sa ukazuje, že oba typy slovných sietí majú *small world* topológiu. Znamená to, že slová nemajú k sebe ďaleko, nech sa na ne

pozrieme z hľadiska sémantického, alebo syntaktického. Preto je mozog schopný v slovníku rýchlo vyhľadávať a komunikácia môže byť plynulá. Štruktúra konceptuálnej siete môže súvisieť s asociačnou pamäťou a aj so štruktúrou neurónovej siete mozgu. Vďaka asociačnej pamäti môžeme pružne nahrádzať jeden výraz iným, podobným výrazom, ktorý sa s ním nachádza v jednom klastri slov. Je to dôležité napríklad vtedy, keď sa chceme vyjadriť v cudzom jazyku a nevie nás správny pojem napadnúť.

Pomocou štruktúry polohovej slovnej siete možno vysvetliť niektoré komunikačné poruchy ako je napríklad *agramatizmus*. Je to porucha, kde mozog nepozná niektoré dobre prepojené slová, ktoré sú kľúčové pre stavbu vety. Tieto slová sa nazývajú funkčnými slovami. V angličtine sú to napr. slová *and, the, of, in...* Majú síce malý sémantický obsah, ale z hľadiska syntaxe sú veľmi dôležité. Zdá sa ako by ich vypadnutie z lexikónu spôsobilo nespojitosť slovného grafu, ktorý síce je odolný voči náhodným poruchám, ale nie voči selektívnemu vymazaniu najlepšie prepojených slov. Ďalšia porucha, nazývaná *paragramatizmus,* obnovuje plynulosť tvorby viet nahradením funkčných slov inými, v danom kontexte neadekvátnymi slovami.

V našom experimente, ktorého výsledky sú zhrnuté na obrázku 2, je podstatné všimnúť si tri veci:

1. Systematickú chybu v škálovacom exponente, tak v štúdii Cancha a Solého [4] ako aj v našej pre strmšiu časť distribučnej funkcie. Hodnota exponentu γ pre túto časť krivky je pod teoreticky predpovedanou hodnotou *3*.
2. Distribučná funkcia zložená z dvoch častí.
3. Iný tvar distribučnej funkcie v oblasti menej strmej časti pre BEV verziu anglického prekladu Biblie.

Tieto tri body je potrebné objasniť:

1. V práci Dorogovtseva a Mendesa [10] je vyriešený model rastúcej siete s preferenčným pripájaním uzlov a s nelinearitou spôsobenou starnutím uzla. Študovaný model je takýto:

$$\frac{\partial k(s,t)}{\partial t} = \frac{k(s,t)}{\int_0^t k(s,t)\, ds}(t-s)^{-\Delta} \tag{15}$$

   Autori v práci [10] dokázali, že pre malú zápornú hodnotu exponentu Δ, γ exponent v distribučnej funkcii je menší ako 3. Čo tento výsledok znamená z hľadiska siete? Znamenalo by to, že nielen počet existujúcich spojení, ale aj vek uzla má vplyv na jeho schopnosť vytvoriť linku s novoprichádzajúcim uzlom.
2. Tvar distribučnej funkcie možno vysvetliť pomocou riešenia rovnice (11). Pre náš experiment to tiež znamená, že buď Biblia neobsahuje poväčšine slová z jadra jazyka, alebo treba hľadať iné vysvetlenie. Text Biblie sa od svojho vzniku výrazne nezmenil. Ostal „zamrznutý“ v pôvodnej podobe, a preto možno odráža situáciu z času svojho vzniku, keď tiež niektoré slová boli z jadra jazyka a iné z periférie.
3. BEV (Basic English Version) verzia anglického prekladu Biblie bola vytvorená pre potreby misionárov a napísaná v silne zjednodušenej angličtine. Preto má

výrazne menší počet slov a veľmi zjednodušenú syntax. Zrejme preto sa BEV krivka na obrázku 2 líši od iných.

Schopnosť slova byť použité vo vete zrejme v priemere súvisí aj s jeho vekom. Ako presne, to si žiada hlbšie skúmanie. Isté však je, že obsah slova sa s časom môže meniť. Predpokladáme, že čím je slovo v slovníku dlhšie, tým lepšia je jeho schopnosť použitia vo vete. Preto predpokladáme, že exponent Δ bude malý a záporný.

Polohový jazykový graf rastie preferenčným spôsobom. To nie je nič zarážajúce, nedá sa predpokladať, že by sa slová radili do vety náhodne. Pre nové slová, teda pre tie, pre ktoré platí $s \approx t$, riešením rovníc (12) a (13) dostaneme $\gamma = 1/2$. Slová, ktoré prišli do lexikónu neskôr, sú horšie prepojené a patria skôr do jazykovej periférie, ako do jadra jazyka. Ak $s \prec\prec t$, slová sú staré a majú šancu mať viacej spojení. Sú to teda slová z jadra jazyka. Predpokladáme, že vzhľadom na body 1. a 2. najlepším modelom pre slovný graf bude asi nejaká kombinácia modelu (11) a (15). Ako bude tento model vyzerať je predmetom ďalšieho skúmania.

Poďakovanie: Táto práca bola podporovaná grantami VEGA NO 1/2045/05, 2/4026/04 a 1/3612/06, 2/7087/27.

Literatúra

[1] Milgram, S.: The Small World Problem. *Psychology Today* **2** (1969) 60-67.
[2] Watts, D. J.: *Small Worlds.* Princeton: Princeton University Press 2004.
[3] Newman, M. E. J.: Models of Small World (review). *J. Stat. Phys.* **101** (2002) 819-841.
[4] Cancho, R. F., Solé, R.: The Small World of Human Language. *Proc. Royal. Soc. London B* **268** (2001) 2261-2265.
[5] Motter, A. E., de Moura, P. S., Ying – Cheng Lai, Dasgupta, P.: Topology of the Conceptual Network of Language. *Phys. Rev. E* **65** (2002) R 065102.
[6] Dorogovtsev, S. N., Mendes, J. F. F.: Language as an Evolving Word Web. *Proc. Royal. Soc. London B* **268** (2001)2603-2606.
[7] Barabási, A. L., Albert, R.: Emergence of Scaling in Random Network, *Science* **286** (1999) 509-512.
[8] Fronczak, A. et al.: Higher Order Clustering Coefficients in Barabási – Albert network, *cond-mat/0212237* (2002).
[9] Orosi, P.: Komunikačná sieť ako graf malého sveta, Fakulta informatiky a informačných technológií STU, záverečný projekt (školiteľka M. Markošová), (2004).
[10] Dorogovtsev, S. N., Mendes, J. F. F.: Scaling properties of scale-free evolving networks: Continuous approach, *cond-mat/0012009* (2002).
[11] http://unbound.biola.edu

Kognitivní aspekty znalostního managementu

Peter Mikulecký[1]

Abstrakt. Cílem tohoto příspěvku je pokusit se zkoumat, zmapovat a přiblížit čtenáři ty aspekty znalostního managementu, na které lze nahlížet jako na kognitivní, uvažovat o nich v relevantních pojmových kategoriích a snažit se najít jistou smysluplnou interpretaci aktivit a přístupů znalostního managementu z pozic kognitivní vědy.

Podnikání je principiálně kognitivní aktivitou. Jinak řečeno, je to aktivita interpretace, porozumění a smysluplnosti.

Michael D. McMaster

1 Úvod

Naše schopnost přetvářet data na informace a posléze na využitelné znalosti může podstatně změnit povahu naší práce, vzdělávání i života. Stále narůstá naše schopnost vytvářet či získávat, modelovat, reprezentovat a aktualizovat složitější a interdisciplinární data a informace z nových a rozmanitých zdrojů. Mohutné možnosti, ukryté v transformaci informací lze však plně zužitkovat pouze pomocí výzkumu, nástrojů a metod nabývání znalostí, klasifikace znalostí, jejich využití a šíření. S tím související teorie a praxe znalostního managementu zaznamenává v současnosti velký rozvoj.

Problematika ***znalostního managementu*** (*knowledge management*) a jeho podpory výkonnými informačními technologiemi je nová a velice aktuální. ***Znalost*** (*knowledge*), kterou z hlediska znalostních systémů můžeme charakterizovat jako ucelenou soustavu poznatků využitelných k řešení problémů určitého typu, se již stala kritickým strategickým zdrojem pro podnikatelské aktivity a další ekonomický rozvoj organizací, a to natolik důležitým strategickým zdrojem, že lze hovořit i o postupném přechodu informační společnosti ve ***společnost znalostní*** (viz např. [1], nebo [2]). Organizace stále více mohou budovat svou konkurenční výhodu na korporativních znalostech, na znalostech zachytávajících nejlepší postupy v organizaci používané, i na znalostech představujících nejnovější pokrok v rozvoji oblasti, ve které působí. K tomu všemu jim je ve vysoké míře nápomocen znalostní management.

1.1 Vymezení znalostního managementu

Znalostní management může, dle četných publikovaných zkušeností (viz např. [1,2,3,4] i jinde), velice zefektivnit práci jak s dokumenty a to zejména s jejich obsahem, tak práci s lidmi. Navíc může velmi přispět ke přirozenému propojení obou těchto součástí. Originální anglický název je „knowledge management“, přičemž překlad tohoto termínu do češtiny není ještě pevně usazen; někteří jej překládají jako „správa znalostí“ nebo dokonce

[1] Katedra informačních technologií, Fakulta informatiky a managementu, Univerzita Hradec Králové, 50003 Hradec Králové, E-mail: Peter.Mikulecky@uhk.cz

„řízení znalostí“. My se přikláníme ke přímočarému překladu do češtiny v podobě termínu „znalostní management“. Uvedeme zde stručné vymezení tohoto pojmu na základě [2], s využitím dalších uvedených pramenů.

Znalostní management lze chápat jako systematický proces vyhledávání, výběru, organizace, koncentrace a prezentace poznatků (znalostí) způsobem, který pomáhá v organizacích zvyšovat úroveň, na jaké zaměstnanci rozumí konkrétním oblastem. Znalostní management tedy pomáhá organizacím docílit hlubšího vhledu a porozumění zejména na základě využití vlastních zkušeností, vlastního intelektuálního jmění. Konkrétní aktivity znalostního managementu pomáhají organizacím zaměřit se na nabývání, uchovávání a využívání znalostí v takových oblastech, jako je řešení problémů, dynamické učení, strategické plánování, rozhodování a další. Ochraňuje také intelektuální jmění organizace před zánikem, přispívá k firemní inteligenci a poskytuje organizaci větší flexibilitu.

Znalostní management není tedy technologií samou o sobě, ani souborem nejlepších postupů, kterých se organizace může držet a lehce je ve své praxi implementovat. Znalostní management je - zajisté kromě svých technologických aspektů - také stavem mysli, tedy přístupem, který se musí rozšířit v rámci celé organizace, má-li být úspěšný. Tento přístup musí pochopitelně zahrnovat také kulturu učení a spolupráce mezi jednotlivci, pracovními skupinami, či organizačními jednotkami organizace (viz např. [3] nebo [4]). Znamená to, že znalostní management se bytostně dotýká lidí a procesů, v rámci kterých lidé sdílejí informace, budují na jejich základě znalosti a přispívají k rozvoji společné, sdílené firemní znalosti.

1.2 Od umělé inteligence ke znalostního managementu

Jak jsme již uvedli, znalostní management můžeme specifikovat také jako souhrn formálních postupů práce se znalostmi, které umožňují jejich vytváření, zpřístupnění a opakované využití, obvykle za podpory moderních informačních a znalostních technologií. Systémy pro podporu znalostního managementu mohou obsahovat početné báze znalostí, sestávající z numerických i kvalitativních dat a z různých typů poznatků. Navíc, systémy pro podporu znalostního managementu často v sobě obsahují také podporu diskusních skupin, které se zaměřují na diskusi konceptů relevantních oblastem které tyto systémy pokrývají.

Existuje celá řada podobností mezi umělou inteligencí a znalostním managementem (viz např. [5]). Lze obecně konstatovat, že znalostní management v mnohém na výzkum umělé inteligence navazuje a využívá její postupy a nástroje. Kupříkladu, systémy pro podporu znalostního managementu využívají bází znalostí, které pak mohou využít lidští uživatelé, ale také jiné počítačové systémy. Postupy získávání znalostí od lidských expertů široce využívají zkušeností již dlouhou dobu rozvíjených a zdokonalovaných v oblasti znalostního inženýrství (blíže viz [5]). V moderních přístupech znalostního managementu báze znalostí obvykle uchovávají vlastní obsah systémů pro podporu znalostního managementu.

Je možno vysledovat několik zásadních typů znalostních bází, využívaných v systémech znalostního managementu. Konkrétní typ obvykle silně závisí od specifik oblasti, ve které daná firma, která využívá daný systém, působí. Kupř. konzultační firmy používají báze znalostí, zaměřené na podporu různých typů odborných činností, ke kterým firma poskytuje konzultace. Běžné jsou také báze znalostí, obsahující přehled novinek

z dané, firmou sledované oblasti. Báze znalostí o nejlepších postupech (*best practice*) poskytují přístup k těm podnikovým procesům, které jsou již vyzkoušenými nejlepšími postupy jak provádět různé činnosti v organizaci (např. výrobní postupy, podnikové procedury, atd.). Báze znalostí o expertech identifikují, kdo je v organizaci expertem pro konkrétní oblast činností a jak je lze kontaktovat.

Významnou součástí znalostního managementu je vytváření a využívání ontologií, zejména pro účely sdílení znalostí. Ontologie se používají ke specifikaci konceptů a relací v rámci dané domény. Lze je využít např. při klasifikaci, modelování systémů, interakci člověka s počítačem, formálním uvažování, či vytěžování dat (*data mining*) v různých aplikačních oblastech, jako je lékařství, fyzika, molekulární biologie, či elektronika.Vývoj ontologií je oblast mimořádně finančně i odborně náročná a proto se jeví v současnosti být jedním z nejdůležitějších problémů souvisejících s oblastí znalostního managementu.

2 Kognitivní aspekty

2.1 Základní východiska a pojmy

O kognitivní vědě se můžeme dovědět více např. v [6]. Pro naše účely použijeme několik pojmů tam zavedených, které pro lepší pochopení textu budeme podrobněji citovat.

Gál [6] uvádí (dle [7]) toto vymezení pojmu ***kognice***:

„Kognice“ odkazuje na všechny procesy, kterými jsou smyslové vstupy transformované, zpracovávané, ukládané, vyvolávané a používané. Týká se rovněž takových procesů, které probíhají za nepřítomnosti relevantních stimulací, jako je kupř. imaginace či halucinace.

Dále jsou v [6] uvedena další upřesnění:

Lidská mysl je tvořená souborem kognitivních orgánů, které jsou specializovány na odlišné perceptuální, intelektuální a behaviorální činnosti, podobným způsobem, jako je lidské tělo souborem orgánů specializovaných na různé fyzické činnosti. Dle Pinkera [8] je lidská mysl systémem výpočetních orgánů, vytvořených přírodním výběrem k řešení problémů, které naši předkové museli řešit v evoluční prehistorii našeho druhu.

Tady se nám nabízí docela zajímavá a celkem přímočará analogie v oblasti organizací, využívajících postupů znalostního managementu: manažerské metody, „leadership“, metody rozhodování, motivační systémy, atd., to vše musely organizace řešit i kdysi, v období „prehistorie“ organizací. Evoluce pak způsobila vývoj metod a postupů moderních a pokročilejších, pomáhajících organizacím přežít ve stále složitějším globálním světě, v podmínkách narůstající konkurence. Bylo tedy přirozené, že „přírodní“ výběr způsobil, že bylo nutno vyvíjet stále nové a nové metody, postupy, způsoby, vyvíjet nové „výpočetní orgány“ organizací, až k dnešním zárodkům „systémů výpočetních orgánů“, neboli ke složitým informačním systémům, systémům pro podporu rozhodování, atd., vesměs k systémům silně ovlivněným umělou inteligencí a vykazujícím její zřetelné prvky.

Tak, jako biologové používají antropomorfizmy, mluví-li o molekulích či bakteriích, nebo tak jako lidé z umělé inteligence užívají antropomorfizmů, mluví-li o robotech, kteří mají preference či plány, o neuronových sítích, které se učí, lze užít (a také se i užívají) docela přirozeně některé antropomorfizmy v oblasti znalostního managementu. Už sám název ***učící se organizace***, často užívaný v této oblasti, je toho pěkným příkladem.

Pokusíme se tedy v dalších částech přinést několik argumentů prezentujících možné pohledy na znalostní management ze strany kognitivní vědy.

2.2 Klíčové aplikace znalostního managementu

Budeme-li pohlížet na zásadní roli znalostního managementu jako na souhrn postupů a přístupů, které v organizaci vedou k takovému sdílení znalostí v rámci organizace, které umožňuje každému jednotlivci či týmu v organizaci je pochopit do dostatečné hloubky a v dostatečném kontextu, aby je mohli efektivně aplikovat při svém rozhodování a zavádění inovací, pak např. dle [3] lze vymezit čtyři klíčové aplikace znalostního managementu. Jsou to:

- zprostředkování (intermediace)
- externalizace
- internalizace
- kognice

Na základě úvah uvedených v [3] se pokusíme předchozí čtyři aplikace podrobněji rozebrat.

Zprostředkování (neboli ***intermediace***) znamená najít a ustavit propojení mezi znalostmi a lidskými uživateli. V lidské společnosti to obvykle znamená, že člověk, hledající konkrétní poznatky, by se měl sejít s jiným člověkem, který je tyto poznatky schopen poskytnout. Tato celkem srozumitelná formulace v sobě skrývá několik úskalí. Předně, člověk hledající konkrétní poznatky nemusí vědět (a velice často to neví), kdo je tím možným poskytovatelem hledaných poznatků. Nebo dále, ví-li to, může se často stát, že v dané chvíli (a často ještě dlouhý čas po ní) nelze z různých příčin s tímto poskytovatelem poznatků hovořit, nebo nějakým jiným způsobem poznatky od něj získat. Lze namítnout, že existují knihy a jiné vhodné zdroje znalostí, ze kterých lze čerpat, nicméně každý, kdo s tím má zkušenosti jistě chápe přednosti toho přístupu, kdy něco důležitého, co rychle chceme použít, nám stručně a jasně vyloží expert na danou problematiku.

Zde tedy sehrává svou významnou roli sdílení znalostí v rámci organizace tak, aby byly přístupny ve vhodné formě každému pracovníkovi, který je může potřebovat. Běžné jsou dva typy zprostředkování znalostí:

- ***Asynchronní zprostředkování*** nastává, když se externalizace a internalizace (viz dále) neobjeví současně. V tomto případě může existovat externí repositář znalostí, poskytující informace o vlastnících potřebných znalostí. Najde-li se při prohledávání báze znalostí potřebná v ní reprezentovaná znalost, nastává internalizace, neboli osvojení si této znalosti tím, kdo ji potřebuje a právě ji nalezl. Je-li však identifikován vlastník tzv. skryté (tacitní) znalosti, internalizace nemůže nastat a dostáváme se právě k případu asynchronního zprostředkování, při kterém musí nejdříve identifikovaný vlastník svou znalost explicifikovat, učinit viditelnou a reprezentovatelnou, a až pak může dojít k jejímu osvojení (internalizaci) zájemcem o její použití.
- ***Synchronní zprostředkování*** nastává vždy tehdy, dochází-li současně k externalizaci i internalizaci. V tomto případě znalost není při svém přenosu uložena v žádné bázi znalostí, ale dochází k přímé komunikaci poskytovatele znalosti a zájemce o její využití. Tento způsob transferu znalostí je velice efektivní v případě skrytých (tacitních) znalostí, známé příklady poukazují na neformální skupiny zájemců diskutujících o předmětu společného zájmu (tzv. communities of practice).

Externalizace znamená propojení znalostí do kontextu s jinými znalostmi. Jejím základem je proces zachycení (zviditelnění a reprezentace) znalosti do nějakého externího repositáře a organizace těchto znalostí dle vhodného klasifikačního rámce či ontologie. Účelem externalizace znalostí je v konečném důsledku poskytnout jejich sdílení.

Uvádí se, že externalizace má obvykle následující tři složky [3]:

- zachycení a uložení znalostí do vhodného repositáře,
- převedení znalostí do podoby, která je využitelná,
- klasifikace nebo uspořádání zachycených znalostí do struktury, která umožňuje jejich efektivní využití.

Internalizace je propojení dotazu, kterého řešení vyžaduje konkrétní znalosti, s relevantní znalostí. Existují dva rozdílné aspekty internalizace:

- extrakce znalostí z repositáře a jejich přenos do mysli člověka, který je potřebuje,
- filtrace znalostí.

Zatím co první aspekt zdůrazňuje nutnost osvojení si zviditelněných znalostí, které byly extrahovány od konkrétních expertů, a reprezentovány v repositáři znalostí, kde ve vhodné formě čekají na svoje využití člověkem, který je potřebuje nebo může potřebovat, druhý aspekt je zaměřen na potřebu vybrat z množství znalostí, které jsou k dispozici, ty „pravé“. Tedy znalosti, které jsou relevantní řešenému problému a potenciálně mohou přispět k jeho vyřešení. Jednou z funkcí postupů znalostního managementu je tudíž i pomoc při filtraci znalostí.

Kognice, jako čtvrtá klíčová aplikace znalostního managementu spočívá v propojení znalostí a procesů, a sice procesů tvorby rozhodnutí založených na dostupných znalostech. Kognice je v tomto smyslu aplikace znalostí, které prošly transferem použitím předchozích tří aplikací, tedy externalizací, internalizací a zprostředkováním (intermediací). Kognice poskytuje uživatelům schopnost vytvářet rychlá a spontánní spojení mezi původně oddělenými poznatky např. použitím visuálních nástrojů umožňujících navigaci v rozsáhlé bázi poznatků. Výsledkem je vytvoření nových poznatků, což je finálním cílem použití postupů znalostního managementu.

V této souvislosti nelze než souhlasit s názorem Kolopoulose a Frappaola [3], že v protikladu s poměrně populárním názorem nemůžeme technologii považovat za všelék pro podporu kognice. Mnohá technologická řešení jsou totiž aplikována na ohraničené problémy, kde prognostika budoucího vývoje je obvykle založena na historických datech a jejich analýze. Problémy a příležitosti pro současné organizace jsou však stále více „nespojité“ povahy a tedy prediktabilita jejich budoucnosti není či nemusí být velmi spolehlivá. Stroje toto tedy zřejmě nedokáží, pouze člověk je schopen myslet mimo zaběhlé škatulky myšlení a dosáhnout použitelných výsledků. Kognitivní přístupy, jako je třeba vizualizace, mohou však sehrát velmi důležitou roli při podpoře lidského chápání kontextu ve kterém má člověk hledat rozhodnutí při řešení problémů.

2.3 Organizace jako kognitivní systém

Velice zajímavé úvahy vedou Vicari a Troilo [9] v rámci podnětného sborníku [4]. Pohlížejí totiž na organizaci jako na ***autopoietický kognitivní systém***. Pokusíme se stručně přiblížit a vysvětlit tyto úvahy.

Prvním předpokladem je, že organizace (nebo firma, chcete-li), je podobně, jako jiné sociální systémy, kognitivní systém. Tato myšlenka poněkud sahá do minulosti, kdy do oblasti ekonomie byl zaveden pojem informace, odkud posléze pramenil mýtus o firmě jako jisté alternativě k individuu zpracovávajícímu informace. V tomto pojetí bylo na firmu ve srovnání s individuem pohlíženo jako na mnohem efektivnější a výkonnější subjekt zpracování informací. Z toho posléze pramenil silný důraz na rozhodovací procesy individua, přičemž byla odmítnuta teze o jeho absolutní racionalitě (později byly pochyby o samém konceptu racionality, ale tyto myšlenky nebudeme dále rozvíjet, neboť by nás zavedly mnohém dále, než máme v úmyslu). Podstatné pro naše úvahy je, že postupně byly formulovány hypotézy o tom, že chování firmy není racionální, ale spíše je vedeno emotivními a intuitivními hledisky (např. práce Mintzberga, nebo Taggarta a Roberyho).

Později, po uvědomení si stále narůstající složitosti rozhodovacích procesů v kontextu organizace, byly vzaty v potaz procesy sociální interakce v rámci organizace a zkoumalo se, jak ony ovlivňují toky rozhodnutí. Posléze, k pochopení fungování organizace bylo nutné zkoumat nejenom procesy individuálního rozhodování, ale rovněž i to, jak sociální systém tvoří rozhodnutí a jak se chová z pohledu interakcí mezi jeho členy. Odtud byl pak pouze krůček k nahlížení na organizace jako kognitivní systémy ve vztahu ke své organizační kultuře, která je dána sdílením norem dané skupiny ze strany jejích členů [9].

V současnosti nastal posun soustředění výzkumu od informace ke znalosti, přičemž je snaha zdůraznit roli znalostí jako základního prvku ve fungování organizace, a tvorbu znalostí jako proces, kterým organizace zvyšuje svou konkurenceschopnost.

Druhým předpokladem se stala skutečnost, že firma je kognitivní systém specifického typu, schopný produkovat své vlastní složky ve spojitém procesu sebe-produkce a sebe-reprodukce, je tedy schopný vytvářet svou vlastní organizaci, i když se v jeho okolí vyskytují perturbace (viz [10]). Firma je tedy autopoietický kognitivní systém, což otvírá celou paletu možností jejího zkoumání z pohledu kognitivních věd. Z tohoto hlediska jsou např. zkoumány vztahy mezi vytvářením nových poznatků a tvořivostí firmy v již zmíněném článku [9].

3 Závěry a důsledky

Cílem tohoto článku bylo pokusit se pohlédnout na moderní a dynamicky se rozvíjející oblast znalostního management očima kognitivních věd. Prezentované úvahy jsou povětšině spíše filozofického charakteru, představují nicméně zajímavé podněty pro jejich další rozvíjení. Možné výsledky a důsledky naznačeného směru uvažování se zdají být slibné, některé zajímavější dosavadní výsledky v tomto směru, dosažené kolektivem autora článku jsou prezentovány v pracích [11] až [20].

Znalost, která se přestala vyvíjet, se mění v pouhý názor nebo v dogma.

Thomas Davenport a Laurence Prusak

Poděkovaní: Tento příspěvek vznikl za podpory Grantové agentury České republiky v rámci grantové úlohy GAČR 406/04/2140 *„KNOMEDIAS – Management znalostí při podpoře aktivit vzdělávacích institucí“*.

Literatura

[1] Davenport, T.H., Prusak, L.: *Working Knowledge: How Organizations Manage What They Know*. Boston, MA: Harvard Business School Press 1998.
[2] Mikulecká, J., Mikulecký, P.: Znalostní management pro informační společnost. *E+M Ekonomie a management* **1** (1998) 41-44.
[3] Koulopoulos, T.M., Frappaolo, C.: *Smart Things to Know about Knowledge Management*. Dover, NH: Capstone US 1999.
[4] Krogh, G. von, Nonaka, I., Nishiguchi, T. (eds.): *Knowledge Creation: A Source of Value*. London: Macmillan Press Ltd. 2000.
[5] Liebowitz, J., Wilcox, L.C. (Eds.): *Knowledge Management and Its Integrative Elements*. Boca Raton: CRC Press 1997.
[6] Gál, E.: *Čo sú kognitívne vedy I.* 2000.
[7] Niesser, U.: *Cognitive psychology.* New York: Appleton-Century-Crafts 1966.
[8] Pinker, S.: „Two Steves“. www.edge.org 1998.
[9] Vicari, S., Troilo, G.: *Organizational Creativity: A New Perspective from Cognitive Systems Theory*. In [6], 63-88.
[10] Maturana, H., Varela, F.J.: *Autopoiesis and Cognition: the Realization of Living.* London: Reidl 1980.
[11] Bureš, V.: Cultural Barriers in Knowledge Sharing, *E+M Ekonomie a Management* **6** (2003) 57-62.
[12] Bureš, V.: Klasifikace znalostí, znalostní strategie a styly znalostního managementu. *E+M Ekonomie a Management* **7** (2004) 43-48.
[13] Bureš, V.: KM-Beat-It: Metodika zavádění znalostního managementu. *E+M Ekonomie a Management* **8** (2005) 36-50.
[14] Gavalec, M., Mls, K.: Individual Cognitive Maps and Evaluation of their Dynamic Changes. In: *Sb. 4. ročníku konference Znalosti 2005 (L. Popelínský, M. Krátký eds.)*, Ostrava: VŠB-TU Ostrava, 2005, 93-96.
[15] Mikulecký, P.: Knowledge Management for Increasing Educational Institution Competitiveness. *E+M Economics and Management* **8** (2005) 106-115.
[16] Mls, K.: From Concept Mapping to Qualitative Modeling in Cognitive Research. In: *Proc. of the First International Conference on Concept Mapping - Concept Maps: Theory, Methodology, Technology, vol. 2, (A.J Canas, J.D. Novak, F.M. Gonzáles eds.)*, Pamplona: Universidad Pública de Navarra, 2004, 271-274.
[17] Mls, K.: Dynamické znalosti jako vzorce v datech a jejich modelování adaptivními fuzzy kognitivními mapami. In: *Sb. posterů 4. ročníku konference Znalosti 2005, (L. Popelínský, M. Krátký eds.),* Otrava: VŠB-TU Ostrava, 2005, 93-96.
[18] Olševičová, K.: Application of Topic Maps in E-learning Environment. In *ITiCSE 05, Proc. of the 10th Annual ACM SIGCSE Conf. on Innovation and Technology in Computer Science Education*. New York: ACM, 2005, 363.

[19] Olševičová, K.: Rebuilding Virtual Study Environments Using Topic Maps. In L. Maicher and J. Park (Eds.): *Charting the Topic Maps Research and Applications Landscape: First International Workshop on Topic Map Research and Applications, TMRA 2005*, New York: Springer, LNAI 3873, 2006, 160 – 168.

[20] Ponce, D.: K časové granularitě poznávání a reprezentace skutečnosti. In *Kognice 2005.* Hradec Králové: Gaudeamus, 2005, 107-113.

Zložitosť (kybernetický pohľad)

Ján Sarnovský[1]

Abstrakt. Cieľom príspevku je poukázať na niektoré aspekty zložitosti z kybernetického hľadiska. Zdôrazňujú sa najmä aspekty riadenia, vzťahu umelej inteligencie, filozofie a zložitosti. Je poukázané na istú spojitosť s klasickými predstavami o systémoch u Hegla. Je pojednané o komplexných adaptívnych systémoch a evolučnej kybernetiky.

1 Úvod

Kybernetika ako veda o riadení zložitých systémov skúma a hľadá niektoré spoločné zákonitosti činnosti strojov, živých organizmov, spoločnosti. Tým aj zdôrazňuje skutočnosť, že neexistuje absolútny rozdiel medzi živou a neživou prírodou, medzi organickým a anorganickým svetom. Pre kybernetiku je typický tzv. systémový, funkcionálny prístup ku skúmaniu problémov a javov okolitého sveta. Tento je tesne zviazaný so základnými pojmami a princípmi kybernetiky. V ďalšom poukážeme na niektoré súvislosti kybernetiky a zložitosti (complexity) ako jednej z dominantnej paradigmy súčasnej vedy.

2 Stručná história kybernetiky

Dejiny kybernetiky majú dve línie. Známa je kybernetika Norberta Wienera, ktorého dokonca označujú za otca kybernetiky. S touto líniou je spojené meno Johna von Neumanna, ktorý mal oveľa širší záber ako Wiener. Ukázalo sa, že kybernetika je oveľa staršia, čo napokon uznal aj Wiener. V novodobých dejinách ho použil známy fyzik Ampére. Tento prevzal termín kybernetika od starých Grékov a sám túto skutočnosť zdôraznil. Slovo kybernetika sa zrodilo na voľnom morskom priestore. Heléni, ktorí prišli na členité brehy Egidy z lesov a stepí indoeurópskej kolísky ľudstva, sa rýchlo stali národom moreplavcov a remeslo kormidelníka, ktorý viedol loď po mori ku ďalekému cieľu, sa tešilo veľkej vážnosti. Homér hovorí o kormidelníkovi kráľa Menelaa Frontisovi, ktorý bol veľmi skúsený a ktorý „lepšie ako ostatní prenikol do tajomstva ako riadiť loď keď hrozí búrka“ a o Faiakoch, ktorí zachránili Odyssea, a ktorých lode samy chápali myšlienky kormidelníkov a zahalené súmrakom a hmlou rýchlo zdolávali more. Aténčania každoročne slávili vo Faleróne kybernézie – sviatky kormidelníkov, ktoré vyhlásil samotný Tézeus na počesť hrdinov Nausitoa (znamená doslova rýchly ako loď) a Faiaxa, ktorí viedli jeho loď na Krétu, kde zabil Minotaura.

Takto sa postupne stalo riadenie lodí názorným a populárnym príkladom a modelom riadenia a slovo „**kybernet**“ **(kybernétes)**, označovalo nielen kormidelníka, ale aj vládcu nad ľuďmi v najširšom zmysle slova, napríklad kráľa, tyrana, archonta, ap. Podobne sa rozšíril aj význam slova „**kybernáo**“, riadim loď a výraz „**kybernetiké téchne**“, umenie riadiť loď, kormidelnícke umenie [8].

Na tieto idey sa odvolával aj Ampére. Grécke *kybernétes* nadobudlo v Ríme podobu *gubernator*. Odtiaľ je aj anglické governor, francúzske gouverneur, rusky

[1] Katedra kybernetiky a umelej inteligencie, FEI TU Košice, E-mail: Jan.Sarnovsky@tuke.sk

gubernátor a po slovensky guvernér. Analogicky z gréckeho *kybernéo* vzniká napríklad sloveso *govern*. Vidíme, že súčasný politický slovník týkajúci sa vládnutia má zdroj v antike. Latinské gubernator na začiatku tiež znamenalo kormidelník. Otázkou je, prečo sa Ampérov metodologický pokus širšie neujal, prečo do zrodu novovekej Wienerovej kybernetiky muselo uplynúť ďalších sto rokov. Odpoveď nie je jednoduchá a už vôbec nemožno pripisovať vinu Ampérovej klasifikácii. Je to problém histórie a metodológie vedy. Jedným z dôvodov je však istá „predčasnosť" Ampérovho pokusu. Problémy, ktoré Ampére vyniesol na svet, ešte nerezonovali ani vo vedeckej komunite ani v spoločnosti. Nemožno nespomenúť ešte pokus Bogdanova (viď jeho tektológiu) na počiatku 20. storočia, ktorý podobne ako Ampére vo svojej dobe ostal nepochopený a jeho prínos, ktorým predbehol teóriu systémov a kybernetiku o vyše dvadsať rokov, bol docenený až v posledných dvadsiatich rokoch.

Problémy riadenia zložitých systémov a s ním spojenej komunikácie sa ešte len začali vynárať. Ako píše Paul Johnson, práve v rokoch 1815-1830 sa rodila nová éra. Pripomeňme si, že v Rusku panoval Mikuláš I., v západných štátoch sa rodilo sociálne hnutie, atď. Problémy s rodiacou sa zložitosťou sveta ako globálneho celku začali až po praktickom využití telegrafu a železnice, čo bolo o pár desaťročí neskôr. Problém bol aj v odlíšení energetických procesov v strojoch od ich informačnej stránky. Práve zásluhou Maxwella, slováka Aurela Stodolu, Ljapunova a ďalších vedcov z oblasti prírodných a technických vied sa v poslednej tretine 19. storočia začína rozvíjať, prakticky aj teoreticky, náuka o regulácii a stabilite strojov a zariadení (napríklad parných turbín). Práve táto linka spolu s rozvojom elektrotechniky, rádiotechniky, biológie a v neposlednom rade aj s projektmi v oblasti vojenstva v súvislosti s 2. svetovou vojnou, viedla v tridsiatych a štyridsiatych rokoch 20. storočia ku vzniku modernej kybernetiky. Ale to sa čitateľ môže dnes dozvedieť z pôvodných Wienerových prác ako aj z početných zdrojov histórie kybernetiky na webe.

3 Umelá inteligencia a filozofia

Je zaujímavé sledovať ako sa rozvíjali predstavy o umelom intelekte a v akej miere je jeho evolúcia porovnateľná s vývojom filozofického myslenia o povahe a štruktúre ľudského rozumu. V samotných začiatkoch rozvoja kybernetiky, v štyridsiatych rokoch, sa vynorili dve predstavy o intelekte. Jeden z nich presadzoval názor, že pri modelovaní rozumu treba vytvárať zariadenie, ktoré bude schopné samoučenia. Tento smer môžeme nazvať neurobiologický. Kládol si za cieľ modelovať psychofyziologické vlastnosti mozgu. Druhý smer, ktorý môžeme nazvať pragmatický, v podstate nezaujímala štruktúra mozgu, ale presadzoval sa model umelého rozumu v podobe programu pre počítač, so systémom povelov, ktoré sa vkladajú zvonku a vopred určujú kroky stroja. Pritom programy riešia „problémy" spôsobom možno úplne odlišným od ľudského myslenia; zaujíma nás však iba konečný výsledok, nie cesta k nemu. Ide teda o funkčný prístup. Už pri nie veľmi hlbokej analýze zistíme, že tieto spory sú veľmi podobné sporom medzi *Lockom* a *Leibnizom*. Locke chápal rozum ako čistú tabuľu, na ktorej sa pocity tak ako postupovali od zmyslových orgánov, zaznamenávajú vo forme „myšlienok". Locke čiastočne ignoroval jazyk ako prostriedok na programovanie činnosti rozumu. Podľa neho sú znakmi myšlienok slová a každý rozum má svoje vlastné myšlienky, čo spôsobuje ťažkosti pri

chápaní. Úloha jazyka je čisto symbolická; je to len súbor dohovorených znakov, nenesie v sebe poznávacie štruktúry.

Leibniz naopak nepovažoval rozum za čistú tabuľu, ale ako miesto na ukladanie poznatkov, ako na „banku údajov, databázu“, kde sa podľa potreby aktivizujú príslušné špecializované programy. Z toho vyplýva aj iný názor Leibniza na úlohu jazyka ako aj jeho snahy vytvoriť presný logický jazyk, ktorý by odrážal univerzálnu štruktúru poznatkov o svete. Lockove myšlienky sa (nezávisle od nich) objavili v kybernetike asi pred 40. rokmi vo forme tzv. rozpoznávania obrazcov. Na základe týchto princípov sa postavili mnohé kybernetické zariadenia, napr. známy *perceptrón.* Tieto boli schopné do istej miery učiť sa na nejakej trénovacej množine, a potom rozpoznávať obrazce. Leibnizova línia (zase nezávisle na jeho ideách) umelého intelektu umožnila vyvinúť programy, ktoré boli schopné porozumieť prirodzenému jazyku, dokazovať teorémy ap. Neskôr sa v tomto smere objavilo aj „heuristické programovanie“. Vedomosti nevyhnutné na riešenie konkrétnych úloh sa vykladali ako súhrn rôznych „heuristík“.

Tento spor mal pokračovanie. D. Hume prisúdil jednej z heuristík – príčinnosti – osobitný štatút. Môžeme ju považovať za akúsi „univerzálnu vysvetľujúcu schému“ na základe zákona dostatočného dôvodu: nemožno si vysvetliť udalosť, ktorá by sa nedala vysvetliť pomocou kategórie príčinnosti. Radikálny záver Huma, ktorý zavrhol možnosť využitia príčinnosti ako logického pojmu, bol rozvinutím Lockových názorov na povahu rozumu, podľa ktorých sa nepripúšťalo apriórne poznanie. Ako vieme opačný názor mal Kant, ktorý odôvodnil existenciu „apriórnych syntetických“ pojmov, ako napr. pojem príčinnosti, priestoru a času. Skúsenosti získané pri rozvoji umelého intelektu ponúkajú (do istej miery) experimentálne overiť tieto filozofické názory na povahu ľudského rozumu.

Na základe posúdenia výsledkov sa dá konštatovať, že teória čistej tabule nebola celkom opodstatnená. Zistilo sa, že možnosti rozpoznávacích zariadení postavených na tomto princípe sú ohraničené. V posledných rokoch sa ukazuje, že do programov umelého intelektu je načim zavádzať „*univerzálne vysvetľujúce schémy*“ (UVS), či „*kognitívne klišé*“. Táto cesta vedie na vytváranie zložitých *neurónových sietí*, čo znamená istý návrat ku neurobiologickým koreňom UI. Takto sa experimentálne rieši polemika medzi Humom a Kantom. V tejto súvislosti je načim skúmať vznik týchto univerzálnych schém. V tomto bode sa teória UI spája so sociológiou v sociálnej oblasti, resp. všeobecnejšie s gnozeológiou (viď. napr. UVS v obdobiach revolúcií, alebo v prípade evanjelií). Proces vzniku UVS nie je zatiaľ podrobne preskúmaný; zdá sa, že dôležitú úlohu tu majú *metafory* s príslušnými objasňujúcimi príkladmi. Metafora plus príklad, to je štruktúra paraboly; táto posilnená mnohými príkladmi sa stáva UVS. Takto sa realizuje prepojenie medzi filozofiou, sociológiou a teóriou umelého intelektu.

Filozofia môže pomôcť riešiť aj niektoré problémy, ktoré vznikajú pri rozpoznávaní reči, resp. všeobecnejšie pri chápaní prirodzeného jazyka. V súčasnosti existujú dva prístupy ku zostaveniu programov riešiacich tieto úlohy. Jeden sa opiera o idey lingvistiky v tom smere, že sa analyzuje slovo rozložením na základné sémantické jednotky a z toho sa reprodukuje zmysel. Druhý prístup je hermeneutický; tu je v každom slove možnosť nekonečnosti významov. Slovo nadobúda zmysel v konkrétnom kontexte, ktorý obmedzuje potenciálnu významovú nekonečnosť. Prvý prístup síce priniesol niektoré pozoruhodné výsledky najmä pri rozpoznávaní slov, avšak zlyháva pri skúmaní

textu ako celku. Umelý intelekt takto vytvára most medzi dvomi kultúrami: humanitnou a prírodovedno – technickou.

Kybernetika ako veda o riadení zložitých systémov vychádza z princípov, ktoré boli uvedené, a ktoré sú aj istým zovšeobecnením princípov rôznych vedných oblastí. Pri skúmaní štrukturálne zložitých javov sa ukazuje nutnosť neustále kriticky prehodnocovať aj samotné princípy. Pri kybernetickom prístupe je potenciálne možné skúmať ľudskú civilizáciu a jej vývoj ako neobyčajne zložitý (kybernetický) systém. Ďalšími dôležitými kybernetickými pojmami sú zovšeobecnené pojmy ako stabilizovateľnosť, pozorovateľnosť, riaditeľnosť. Základný princíp homeostázy môže určovať možný cieľ alebo úlohu evolúcie zložitých systémov. Aj také vlastnosti ako sú adaptabilita živého systému vyplývajú z princípu homeostázy. Nevyhnutným znakom inteligencie je predovšetkým schopnosť systému uchovávať a spracovať informáciu. V súlade s predstavami o formách spracovania informácie môžeme pristúpiť ku zložitejšej predstave – ku „*logike systému*“. Je to schopnosť systému odrážať vonkajšie zákonitosti okolitého prostredia (ako takéto prostredie môžu vystupovať samotné časti systému) pomocou určitej množiny vnútorných reakcií (obvykle v kódovanej forme). Tieto sa potom využívajú na analýzu a predikciu vonkajších situácií s cieľom „zlepšiť“ svoju vlastnú existenciu (homeostázu).

V súčasnosti sa skúmajú problémy „foriem logík“ v súvislosti s rozvojom metód UI. Je to napr. „*fuzzy*“ logika. Tieto logiky môžu pomôcť vyjadriť istý stupeň dialektiky, napr. typy záporov starej indickej logiky: niečo niekedy bolo a nemôže byť; niečo bolo, ale teraz to nie je; niečo teraz je, ale skoro to zmizne. Náš intelektuálny svet obsahuje rôzne logiky, čo vytvára rôzne svety (napr. logika absurdity rozprávkového sveta). Vo vnútri každého z nich je logika bez rozporov, ale existujú „značky“ prechodov zo sveta do sveta.

Na základe podobných úvah S. Lem pokladá za možnú cestu ďalšieho vývoja vytvorenie „*sveta vo svete*“, čiže nášho *posthominálneho sveta*, t.j. istej sústavy umelých podmienok, ktorý by sa riadil sústavou programovaných, umelých zákonov tohto „sveta“. Niet pochýb o tom, že podobné hypotézy sú veľmi sporné. Najdôležitejší záver je v tom, že ak sú princípy činnosti zložitých systémov poznateľné, potom rozvoj vedy môže pomôcť určiť nejakú „optimálnu“ cestu vývoja. To vedie ku otázke, aký má byť cieľ ďalšieho vývoja? Cieľom činnosti zložitého systému môže byť pochopenie princípov jeho činnosti. Objektivita kategórií *rozum, vedomie, logika systému* súvisí s funkčnými vlastnosťami systému. To, že sa objavuje istá „metafyzika“, dogmy, paradigmy, ap., je spojené s reálnou činnosťou systému. V danej etape vývoja systému, (sveta, spoločnosti), nedisponujeme úplnou informáciou (poznaním) o okolitom prostredí, avšak chováme sa tak, ako by poznanie bolo úplné. Definitívnosť akčných zásahov je daná presvedčením o ich oprávnenosti a správnosti ako aj o riaditeľnosti systému; inak by žiadna činnosť systému nebola možná. Práve v tom spočíva prirodzená „metafyzika“, v širšom zmysle slova „filozofia“ a „dogmatizmus“. Vo svojom vývoji prekonáva systém celý rad paradigiem, pracovných hypotéz, ktoré určujú jeho praktickú činnosť. Tento proces však môže vyústiť do „extrémneho“ dogmatizmu ak nie je neustále konfrontovaný so skutočnosťou, preverovaný experimentom. Skutočná sila pracovných hypotéz, paradigiem, dogiem, je v ich neustálej premenlivosti, v ich adaptabilite s vedeckými poznatkami.

Riešiť otázky a problémy cieľov civilizácie predpokladá teda podstatný pokrok v teórii kybernetických systémov. Súčasná úroveň ich poznania zatiaľ neumožňuje robiť

definitívnejšie závery. Významné však môžu byť metodologické a najmä filozofické závery, ktoré môže odhaliť skúmanie činnosti neobyčajne zložitých systémov.

4 Zložitosť

Stanislav Lem uvádza mimoriadne zaujímavú klasifikáciu vo svojej knihe Summa technologiae. Rozlišuje niekoľko skupín systémov, pričom vlastnosti nasledujúcej triedy obsahujú v sebe vlastnosti predchádzajúcej. Do prvej skupiny patria deterministické systémy; sú to jednoduché systémy alebo len pracovné nástroje, ktoré nie sú spojené so svojím okolím spätnou väzbou. Do druhej triedy patria organizované samočinne sa regulujúce systémy so spätnou väzbou. Do tejto širokej triedy patria aj automatické systémy, teda systémy s riadenou (regulovanou) dynamikou, ale aj človek a biologické systémy. Tretia skupina zahrňuje systémy, ktoré môžu meniť svoje usporiadanie, aj svoje „princípy“, pričom využívajú materiál, ktorý je k dispozícii. Príkladom je biologická evolúcia. Lem predpokladá aj systém štvrtej skupiny, kde je možnosť vyberať z okolia vhodný materiál a takto sa sebareprodukovať. Za takýto systém možno považovať vedecko-technickú aktivitu ľudstva. Podľa Lema je možné uvažovať ešte o ďalšej skupine, ktorá materiál pre svoju sebareprodukciu nevyberá z hmoty, ktorá existuje vo vesmíre, ani nevyužíva fyzikálne a chemické technológie, ale vytvára umelé, „syntetické“ podmienky, ktoré pri prirodzených fyzikálnych podmienkach nemôžu vzniknúť. Ide o nové formy existencie hmoty, ktoré ľudstvo dokáže v budúcnosti realizovať.

Uvedená, alebo podobná klasifikácia má svoj význam, pretože umožňuje určiť miesto systému, ktorý nás zaujíma, ako aj analyzovať systém z požadovaného hľadiska. Napríklad nás môže zaujímať problém stability (trvanlivosti) systému a pri zložitejších alebo zložitých systémoch zvlášť problém jeho regulovateľnosti (riaditeľnosti, ale nie v úzkom matematickom zmysle). Lem v jednom z jeho románov – Pánov hlas – píše:

„*Pred devätnástimi rokmi som s mladým antropológom uverejnil prácu, v ktorej som zdôvodnil, že existuje prah komplikácií pre dokonalé algedónicky riadené automaty, ku ktorým patria všetky zvieratá spolu s človekom. Algedónické riadenie znamená osciláciu medzi trestom a odmenou – ako bolesťou a rozkošou.*

Môj dôkaz vysvetľuje, že ak počet prvkov regulačného centra (mozgu) na najvyššej úrovni presahuje štyri milióny, súbor takýchto automatov prejavuje rozptyl medzi nezhodnými článkami riadenia. V každom takom automate jeden z kontrolných pólov môže prevažovať, alebo bežnejším jazykom povedané, sadizmu a masochizmu nemožno uniknúť a ich vznik v procese antropogenézy bol nevyhnutný. Evolúcia sa pre takéto riešenie ‚rozhodla‘, pretože sa riadi štatistickým počtom; jej sa ráta zachovanie rodu, a nie chyby, choroby, trápenia – niektorých jednotlivcov. Vývoj je oportunistický konštruktér a nesmeruje k dokonalosti.“

Problém regulovateľnosti súvisí s jedným z najdôležitejších princípov kybernetiky, *princípom homeostázy.* Homeostáza je taký princíp činnosti systému, pri ktorom sa hodnoty niektorých podstatných premenných udržujú v určitých medziach, aj pri pôsobení vonkajších vplyvov.

Pojmy ako zložitosť (complexity), chaos, samoorganizácia, vznik (emergence), sú v súčasnosti v rôznych vedných oblastiach, ale aj vo vedecko – populárnych esejach veľmi často používané. Čo ešte viac priam bije do očí je neadekvátne používanie týchto pojmov. Na WWW môžeme nájsť obrovské množstvo rôznych „centier“ pre výskum zložitých

systémov (compex systems). Pojem zložitosti, zložitých systémov a ich riadenia je ústredným kybernetickým pojmom. Nie nadarmo sa kybernetika jednoducho definuje ako veda o riadení zložitých systémov. W. R. Ashby vo svojej knihe „Kybernetika“, predstavuje koncept systému ako ústredného pojmu modernej vedy. V úvodnej kapitole Ashby píše: „*Veda dnes stojí na akomsi rázcestí. Dve storočia sa skúmali systémy, ktoré sú pomerne jednoduché, lebo môžu byť rozčlenené na jednotlivé zložky. Že sa po stáročia uznávala taká dogma ako ‚meňte činitele jeden po druhom‘, ukazuje, že sa vedci zaoberali v podstate skúmaním takých systémov, ktoré túto metódu umožňovali; v zložitejších systémoch sa však táto metóda nedá často zásadne vôbec použiť.*“

Systém je zložitý, ak je ho možno opísať množstvom rôznych spôsobov, z ktorých každý korešponduje s iným subsystémom. Zložitosť je z latinského *complexus*, čo znamená zvinutý, skrútený spolu. To znamená, že takýto systém musí mať viac prvkov a tie musia byť nejakým spôsobom spojené tak, že je ich ťažko oddeliť. To je skrytá duálna vlastnosť častí, ktoré na jednej strane chápeme ako rozdielne a na druhej strane ako spojené. Z toho vyplýva, že skutočne zložitý systém nemožno analyzovať ako množinu nezávislých častí bez nebezpečia zničenia systému. Obidva krajné prístupy, redukcionizmus a holizmus, nevedú ku cieľu. Potrebujeme prístup, ktorý využíva obidva prístupy. Jednou z možných ciest je chápanie zložitého systému ako *siete (network)*.

Často sa zdôrazňuje fakt, že celok je viac ako prostá suma jeho častí. Problematika celku a častí je vo filozofii veľmi dobre rozpracovaná a pokusy špeciálnych vedcov (myslí sa vedcov z konkrétnych vedných oblastí) k tomu čosi dodať obvykle dopadnú trápne. Pojem častí a celku sa objavuje už v diskusiách starovekých filozofov o vzťahu jediného, jednotného poznávaného bytia k jeho mnohotvárnym, meniacim sa, zmyslami vnímaných foriem. Už Herakleitos odlišoval časť a celok. U Platóna sa výrazne prejavuje myšlienka prvotnosti celku, no celok sa ako idea podrobuje častiam ako cieľ teleologického smerovania. Hlboké myšlienky o vzťahu časti a celku vyslovil Kant, no najmä Hegel. Vzťah častí a celku chápe ako jednotu protikladov. Celok nemôže existovať bez častí a časti ako také nemôžu existovať bez celku, ktorý formuje ich podstatu. U Hegla nachádzame aj návod, dnes by sme povedali, algoritmus, ako túto kontradikciu prekonať. Aby sme pochopili jeho myšlienky, najlepšie je ich citovať:

„*Celok sa teda v častiach rovná iba sebe samému a rovnakosť celku a častí vyjadruje iba tautológiu, že* ***celok*** *sa* ***ako celok nerovná častiam****, ale* ***celku****. A naopak, časti sa rovnajú celku; keďže však samy obsahujú moment inobytia, nerovnajú sa mu ako jednote, ale tak, že* ***jedno*** *z jeho rozmanitých určení pripadá na časť, čiže sa mu rovnajú ako* ***rozmanité****, t. j. rovnajú sa mu ako* ***podelenému celku, teda ako častiam****. Tak vzniká tá istá tautológia, že sa* ***časti ako časti*** *nerovnajú* ***celku ako takému****, ale že sa v ňom rovnajú* ***sebe samým, častiam****. Celok a časti sa takto ľahostajne rozpadávajú; každá z týchto stránok sa vzťahuje iba na seba. Keď sú však takto rozpadnuté, samy sa ničia. Celok, je ľahostajný voči častiam, je* ***abstraktnou****, v sebe nerozdielnou* ***identitou****; táto identita tvorí celok ako* ***rozdielna v sebe****, a to tak, že tieto rozmanité určenia sú reflektované do seba a bezprostredne samostatné.*

Pravdou vzťahu je ***sprostredkovanie****; podstatou vzťahu je negatívna jednota, v ktorej bola prekonaná tak reflektovaná, ako aj jestvujúca bezprostrednosť. Vzťah je protirečenie vracajúce sa do svojho dôvodu, do jednoty, ktorá, keďže sa vracia, je reflektovanou jednotou. ... Keďže si antinómiu tohto úsudku všimneme celkom zblízka, hovorí vlastne toto: Keďže celok nie je samostatný, je samostatná jeho časť; keďže však*

*časť je samostatná iba **bez celku**, nie je samostatná **ako časť**, ale ako **celok**. Nekonečnosť takto vznikajúceho progresu je podmienená neschopnosťou spojiť obidve myšlienky, ktoré obsahuje sprostredkovanie hovoriace, že každé z dvoch určení sa v dôsledku svojej samostatnosti a izolovanosti od druhého stáva nesamostatným a prechádza do druhého. Tak prešiel vzťah celku a častí do vzťahu **sily a jej prejavenia sa navonok.***

***Sila** je negatívna jednota, v ktorej sa rozplynulo protirečenie medzi celkom a časťami, je pravdou onoho vzťahu. Podstatný vzťah sa teraz určil tak, že bezprostredná a reflektovaná samostatnosť sú v sile kladené ako prekonané čiže ako momenty, ktoré boli v predchádzajúcom vzťahu pre seba pretrvávajúcimi stránkami, čiže krajnými termínmi. Reflektovaná jednota, **sila, sa prejavuje navonok**. **Sila** je jednotou bezprostredného a reflektovaného, čiže jednota formy a vonkajšej samostatnosti. Je obidvoje v jednom, je dotykom takých daností, z ktorých jedna je vtedy, keď druhá nie je. Forma sily je podmienená a jej obsah obmedzený; lebo určenosť formy implikuje aj obmedzenie obsahu.*“ [9]

Aj keď uvedený výklad je trochu komplikovaný, ak si dosadíme za **silu** pojem **riadenie, môžeme povedať, že antinómia častí a celku zložitého systému je prekonaná a vyriešená riadením tohto systému.** Samotné riadenie môže už byť súčasťou zložitého systému ako jeho organickej súčasti, alebo je zavedené zvonku. Takto riadenie je tým *glejom*, tým navonok záhadným mystickým prvkom, ktorý je mostom medzi celkom a jeho časťami.

5 Komplexné adaptívne systémy

Komplexnosť (zložitosť, complexity) je dnes jedným z ústredných pojmov súčasnej vedy a ako taká je hlboko študovaná v rôznych oblastiach ako sú fyzika, biológia, sociálne systémy, ap. Len multidisciplinárny výskum môže priniesť konkrétne výsledky, syntézou ktorých môžeme postupne dospieť ku všeobecným princípom a zákonom riadenia komplexnosti. Na druhej strane sú to globálne svetové problémy, napríklad fenomén terorizmu, ktoré sú komplexné a ich riešenie si vyžaduje modelovanie, a potom ich riadenie.

Zdrojom zložitých úloh sú komplexné systémy. Tieto majú obvykle autonómne časti. Ich vzájomné interakcie spôsobujú ťažkosti pri snahe redukovať a zjednodušovať komplexný systém. Práve posudzovanie systému ako celku nám ukáže jeho zložitosť. Rýchly rozvoj nanotechnológie a ich konvergencia s biologickými, informačnými a kognitívnymi vedami vytvára novú oblasť komplexných systémov. Komplexné systémy vychádzajú na jednej strane z detailných výskumov konkrétnych systémov, ako aj na druhej strane zo všeobecných skúmaní opisu a reprezentácií zložitých systémov. Aplikácie v praxi, v biologických, informačných, kognitívnych, sociálnych systémoch ako aj v inžinierskych disciplínach sú zjavné. Biológia zhromaždila obrovské množstvo poznatkov a dnes je jasné, že biologické systémy sú zložité, bohato štruktúrované biochemické siete. Úloha informácie v biologických systémov začínajú študovať matematici, fyzici a inžinieri, ktorí hľadajú v evolučne vytvorenej komplexite vzor funkčnej spoľahlivosti. Počítače prekonali vývoj od jednoduchých individuálnych systémov ku zložitým sieťovým štruktúram.

Hlavnými cieľmi výskumu KS je:

- porozumieť ako fungujú a ako kooperovať s KS v rôznych úrovniach (manažérska, inžinierska, ekonomická, sociálna, ap.)

- porozumieť jednotným princípom organizácie a riadenia KS, najmä systémov, v ktorých sa vyskytujú veľké objemy informácií
- porozumieť komplikovaným interakciám medzi KS a prostredím

Porozumieť KS neznamená to, že dokážeme presne predikovať ich správanie. To nie je len skonštruovanie veľkých databáz a robenie masívnych simulácií; zdá sa, že dôležitejšie je, že sa dozvieme, aké sú hranice nášho poznania, čo môžeme a čo nemôžeme robiť.

Špecifickou triedou KS sú komplexné adaptívne systémy. Pojmy chaos, zložitosť, komplexné adaptívne systémy, sú dnes synonymami, ktoré charakterizujú dnešnú dialektickú zložitosť sveta ako celku. Zároveň sú výrazom istej módnosti a používanie týchto pojmov často nie je na mieste. V našom prirodzenom ako aj umelo vytvorenom svete ide o také systémy ako mozog, ekonomika (podniku, štátu, svetadiela), kolónie mravcov či včiel. Ešte zložitejšie sú sociálno-ekonomicko-kultúrne systémy. Tieto systémy môžeme charakterizovať prostredníctvom ich niektorých podstatných vlastností:

- Komplexné adaptívne systémy (KAS) sú siete vytvorené z mnohých podsystémov – agentov, ktoré pôsobia paralelne. Ak uvažujeme napr. ekonomické systémy, agenty sú firmy. V politickom systéme sú agentmi politické strany, prípadne dominantné osobnosti politickej scény. V medzinárodných vzťahoch sú to národy a štáty, prípadne zoskupenia štátov. V zložitom energetickom systéme sú to jednotlivé energetické jednotky. Typické pritom je, že agent pôsobí v neustále premenlivom prostredí, v ktorom je neustále v interakciách s ostatnými agentmi a to si vyžaduje jeho neustálu adaptáciu.
- Riadenie KAS je decentralizované a autonómne. Napríklad v mozgu nejestvuje žiadny vedúci neurón. Aj v technických systémoch existuje mnoho príkladov autonómnych systémov bez koordinátora. Činnosť systému ako celku je zabezpečovaná kooperáciou alebo konkurenciou medzi agentmi.
- KAS majú obvykle mnohoúrovňovú štruktúru s decentralizáciou na každej úrovni.
- KAS neustále menia svoju štruktúru ako reakciu na vonkajšie prostredie. Pritom princípy adaptácie sú rovnaké na každej úrovni.

6 Evolučná kybernetika

Evolučnú kybernetiku môžeme definovať ako štúdium procesu variácie a selekcie zložitých systémov riadenia. Tento problém môžeme skúmať z dvoch stránok: kvalitatívne, ako vzniká úroveň riadenia v systéme, a kvantitatívnej, t.j. ako sa systémy riadenia vyvíjajú. Evolučný vznik riadenia môžeme nazývať metasystémovou tranzíciou (viď. [5]). Darwinova teória vysvetľuje vznik riadenia: prirodzená selekcia preferuje systémy, ktoré môžu prežiť napriek vonkajším perturbáciám a to sú systémy s riadením. Štúdium takýchto systémov z pohľadu napr. uplatnenia spätnej väzby, nám môže odhaliť skúmané procesy evolúcie z hľadiska riadenia.

Evolučná kybernetika je spätá s novými disciplínami, napr. s evolučnými systémami (viď. systémová teória evolúcie). Kým systémový prístup viac zdôrazňuje štruktúru systémov, kybernetika si viac všíma ich funkčnú a cieľovú stránku. V súčasnosti je veľmi aktuálny a populárny výskum v oblasti komplexných adaptívnych systémov. Tento je

orientovaný na využitie matematických metód, napr. teórie chaosu a pod., ďalej fyziky a chémie. Oveľa viac vplýva paradigma kybernetiky na také oblasti ako sú umelý život, multiagentové systémy, ktoré skúmajú problémy vývoja autonómnych cieľovo orientovaných podsystémov (agentov). Samozrejme, že okrem vytvárania modelov a aplikácií vo vybraných oblastiach je potrebné venovať viac pozornosti teoretickým a filozofickým aspektom a aplikácií v oblasti humanitných vied; viď. napr. rozvoj nových vedných oblastí ako sú evolučná epistomológia, memetika, apod. To umožní multidisciplinárny výskum v smere naznačenom a realizovanom v projekte Principia Cybernetica Project.

Evolučné procesy boli implicitne odjakživa skryté v princípoch a metódach kybernetiky (dnes hovoríme o evolučnej kybernetike). Kybernetika prešla vo svojom vývoji rôznymi etapami, pričom sa postupne strácal jej holistický integrujúci charakter. Celkom zákonite vznikali „nové“ smery, napr. neurónové siete, komplexné adaptívne systémy, či umelý život a samotná kybernetika akoby sa vytrácala. Asi dozrel čas sa vrátiť ku niektorým východiskám kybernetiky ako vedy o riadení zložitých systémov, prehodnotiť a integrovať nové pohľady najmä na zložitosť a evolúciu.

Systém, ktorý obsahuje veľké množstvo takýchto častí môžeme klasifikovať podľa klasika kybernetiky Stafforda Beera ako neobyčajne zložitý [3]. V tomto prípade Bar-Yam navrhuje použiť evolučný prístup [2]. Evolučný proces sa všeobecne chápe ako analógia súťaže v prostredí voľného trhu a vychádza z iteračného procesu postupných zmien. Existujú však zásadné rozdiely medzi evolúciou a procesom postupných zmien používaných v technike. Evolučný prístup vyžaduje existenciu viacerých odlišných systémov a zmeny sa vyskytujú súčasne a paralelne. Paralelné testovanie viacerých rôznych zmien vo viacerých systémoch a ich kombinácia je podstatne odlišná od konvenčných postupných inžinierskych prístupov. Odlišnosť je aj v tom, že proces projektovania vedie nakoniec ku jedinému riešeniu a nie ku paralelnej implementácii. Ďalšia odlišnosť evolúcie je aj v tom, že implementácia a testovanie sa vykonáva takpovediac „v poli“, v procese učenia prostredníctvom spätných väzieb s okolím. Evolúcia je najmä o zmenách v populácii organizmov v čase, pričom rozhodujúcu úlohu má kritérium efektivity (všeobecne známy pojem „fitnes“). Vo všeobecnosti je evolúcia proces replikácie s variáciami a následnou selekciou; to všetko na báze súťaže. V procese inžinierskeho projektovania, riadenia a výroby má však veľkú úlohu aj kooperácia. Taktiež treba odlíšiť pojem evolučného projektovania a riadenia od použitia evolučných algoritmov ako prostriedku výpočtovej inteligencie pri klasickom projektovaní. Kooperácia a konkurencia sa vzájomne nevylučujú, najmä ak sa vyskytujú na rôznych stupňoch organizácie a riadenia.

Riešiť problém na úrovni civilizácie si vyžaduje také informácie a znalosti, ktoré nám umožnia vyvinúť aspoň homomorfný model civilizácie, ktorý umožní pri stanovených cieľoch riadenie civilizácie ako kybernetického systému. Zatiaľ to asi nevieme. Avšak rozvoj vedy nám dáva nádej, že tieto problémy môžeme riešiť. V prvom rade je jasné, že kybernetický systém musí byť homeostat. Homeostáza je základný kybernetický princíp a môžeme ho prirovnať ku zákonom zachovania vo fyzike – všetky premenné systému ostávajú aj pri vnútorných aj vonkajších poruchách v predpísaných intervaloch. Ashby ako biológ zaviedol tento pojem páve podľa vzoru fungovania biologických systémov.

Čo hovorí o riadení zložitých systémov kybernetika? Už klasici kybernetiky Walter Ross Asby a Stafford Beer sformulovali dva veľmi dôležité princípy (niekedy sa

hovorí aj zákony) riadenia veľmi zložitých systémov. Prvý, princíp nevyhnutnej variety hovorí, veľmi jednoducho povedané, že len zložitosť (Ashby ju nazýva varieta) môže „zvládnuť“ zložitosť. Tento princíp sa nazýva aj princíp adekvátnosti systému a regulátora. Len zložitý regulátor môže regulovať zložitý systém. Uvedený zákon sa dá sformulovať aj matematicky a dá sa dokázať. Metodologicky z neho vyplýva dôležitý dôsledok: zložité systémy nemožno riadiť jednoducho. Z uplatnenia tohto princípu tiež vyplýva, že ideálny regulátor (vzhľadom na poruchy z okolia) je takzvaný kompenzačný, alebo dvojkanálový, v otvorenom obvode. To si však vyžaduje izomorfný (jednoducho úplne presný) model systému. To však v praxi nikdy nie je splnené, k dispozícii máme len homomorfný (približný, jednoduchší) model, a preto musíme použiť riadenie so spätnou väzbou. Takže výsledné riadenie zložitých systémov je vždy kombinované. Má zložku cieľovú, programovú (vyjadrením je plán), kompenzačnú (vyžadujúcu si izomorfný model) a zložku so spätnou väzbou.

Druhý princíp, vonkajšieho doplnenia, odvodil Beer z Goedelových teorém. V systéme riadenia a rozhodovania môžu vzniknúť situácie a stavy, ktoré sú na danej úrovni „nerozhodnuteľné“ (napríklad štrukturálne zmeny dynamického systému si vyžadujú okrem spätnej väzby od výstupu doplnkovú spätnú väzbu od odchýlky výstupov želaného-etalónového modelu a výstupu reálnej sústavy). Túto druhú väzbu môžeme chápať ako nadradenú a ako „vonkajšie doplnenie“ do regulátora, a teda ako vyššiu, nadradenú úroveň riadenia. Tento princíp je teda metodologickým základom hierarchických, alebo mnohoúrovňových systémov.

O akú zložitosť ide v sociálno-ekonomických systémoch? Je to zložitosť ohromujúca, no konečná. Závisí aj od formulácie problémov a najmä od vhodného počiatočného obmedzenia variety. Ako príklad uvádza Beer premávku na cestách a križovatky. Jednoduchými pravidlami (je to riadenie v otvorenom obvode bez spätnej väzby) sa hneď na začiatku dosiahne podstatné obmedzenie variety (možných trajektórií áut po ceste). Skvalitnením riadenia, ešte stále v otvorenom obvode, je zelená vlna. Ak namontujeme na semafory snímače počtu áut (teda meriame hustotu premávky), môžeme túto informáciu využiť na spätnoväzobné riadenie (adaptívne zmeny časových intervalov na semaforoch) a vznikne stabilnejší a kvalitnejší, kombinovaný systém riadenia. Zložitosť sociálno-ekonomických systémov je vyjadrená, napríklad, počtom obyvateľstva v určitom regióne, alebo aj na celej zemeguli a interakciami medzi nimi. Aby sme mali nejakú predstavu, moderné dopravné lietadlo má asi 100 tisíc súčiastok, kód informačného a riadiaceho systému takéhoto stroja má asi 20 miliónov riadkov textu. Európska elektrizačná sústava má rádovo desaťtisíce generátorov, navzájom prepojených s mnohými miliónmi spotrebičov, ľudský mozog má desiatky miliárd neurónov a oveľa väčší počet prepojení. Napríklad ľudský genóm má vyše troch miliárd písmen. Rozlúštenie ľudského genómu umožnilo vytvoriť jeho databázu, ktorá je de facto jeho izomorfným, alebo homomorfným modelom. Jedinou šancou ako zvládnuť zložitosť tohto systému je vytvoriť regulátor s takou varietou, aby zvládlo varietu (zložitosť) ekonomiky celého sveta. A dobré si je zopakovať, len varieta môže zničiť varietu. Len zložitosť môže zvládnuť zložitosť. Inak sa svet bude zmietať v neustálych evolučných krízach a bude len reagovať na situáciu. Riadenie so spätnou väzbou je vynikajúce pre technické systémy, kde časové konštanty dynamiky systému sú relatívne krátke – nanajvýš hodiny, ale veľmi často minúty, sekundy, ba desatiny až tisíciny sekundy. Oneskorenie, ktoré vzniká pri regulácii so spätnou väzbou, neovplyvňuje v konečnom dôsledku dobrú činnosť systémov. Inak je to pri riadení spoločensko-

ekonomických systémov: ide o roky, ba desaťročia. V tomto prípade je riadenie so spätnou väzbou neúčinné, pretože nie je možné riadiť systém v reálnom čase.

Riadenie v reálnom čase je jeden zo základných konceptov kybernetiky. Spätná väzba je účinná vtedy, ak časové oneskorenie v uzavretom regulačnom systéme nie je také veľké, že spätná väzba fakticky nejestvuje. Ide o prípady systémov, ktoré sme uviedli v eseji dejiny, súčasnosť a budúcnosť. Ako príklad sme uviedli systémy ako históriu a astronomické objekty, kde sa informácia o objekte nedá využiť pre riadenie v reálnom čase. A keďže spoločensko-ekonomické systémy sú systémy historické, vyžadujú si iné riadenie. Ako to hovorí Ashby, v tomto prípade je nutné použiť aj riadenie cause-control (kompenzačné riadenie), ktoré zabezpečuje potlačenie vonkajších vplyvov okolia. Takéto riadenie nemá oneskorenie, avšak vyžaduje si veľmi presný model riadenej sústavy, teda izomorfný, alebo homomorfný, ale so značným priblížením ku izomorfizmu. To je hlavný dôvod, prečo musíme v ekonómii vytvoriť zložité modely ekonomickej reality, napríklad na úrovni ľudského genómu v biológii. Teda nie nekonečné hračky s dilemou banditov a s modelmi iných primitívnych hier, ale skutočný model s miliónmi premenných. Vytvorenie takýchto zložitých modelov a ich použitie pre riadenie ekonomiky je krok ku premene ekonómie na exaktnú vedu.

Poďakovanie: Tento príspevok vznikol za podpory grantovej agentúry Vega v rámci grantovej úlohy 1/2183/05.

Literatúra

[1] Ashby, W. R.: *Kybernetika*. Praha: Orbis 1961.
[2] Bar-Yam, Y.: *Enlightened Evolutionary Engineering/Implementation of Innovation in FORCEnet,* Report to Chief of Naval Operations Strategic Studies Group, 2002 (Brief 2000).
[3] Beer, S.: *Kybernetika a řízení*. Praha: Svoboda 1966.
[4] *History of Cybernetics,* http://www.asc-cybernetics.org
[5] *Principia Cybernetica Poject,* http://pespmc1.vub.ac.be
[6] Sarnovský, J.: *Kybernetický svet.* Košice: Elfa, 2. vydanie, 2002.
[7] Wiener, N.: *Kybernetika a společnost*. Praha: Nakladatelství ČSAV 1963.
[8] Povarov, G.N.: *Ampére a kybernetika*. Praha: SNTL 1980.
[9] Hegel, G. F. W.: *Logika ako veda, I., II.* Bratislava: Pravda 1985.

Myseľ, inteligencia a inteligentné technológie

Pohľad obyčajného učiteľa na inteligentné technológie

Peter Sinčák[1]

Abstrakt. Informačné systémy dnes prenikajú do všetkých oblastí technickej aj netechnickej praxe. Dnes bez informatizácie a nasadenia informačných systémov v podstate neexistujú žiadne technológie. Ich úroveň sa neustále zvyšuje s cieľom vytvorenia používateľsky príjemných systémov, ktoré zvládajú mnohé situácie, ktoré je možné zvládnuť bez zásahu človeka. Resp. sa realizujú tak, ako by ich riešil človek s určitými vedomosťami, ktoré má tento systém zachytené vo svojej báze znalostí, ktorú neustále zväčšuje. Informačný systém akoby „dospieval" a poskytuje pre človeka inteligentnú podporu v jeho práci a rozhodovacích procesoch. Takéto systémy nazývame tzv. učiace sa systémy a v poslednej dobe tzv. systémy so strojovou inteligenciou alebo tzv. inteligentné technológie.

1 Čo je to strojová inteligencia?

Keď ma prof. Kvasnička požiadal o napísanie príspevku do tejto monografie samozrejme som súhlasil, ale postupne som si začal uvedomovať, že možno sklamem niektorých čitateľov. Témy ako myseľ, inteligencia majú veľmi dôležitý vedecký rozmer a je veľmi dôležité tieto veci študovať po teoretickej ako aj experimentálnej línii. Moja životná ambícia je trochu iná a smeruje ku malému príspevku presadzovania inteligentných technológií v slovenských a európskych podmienkach. Tieto sú veľmi dôležité, lebo posúvajú inovačný faktor ľudského bytia a vedia napomáhať zvyšovaniu kvality života v spoločnosti. Preto sa o to snažím a verím, že sa mi podarí inšpirovať niekoľko ľudí (mojich študentov) k zavádzaniu týchto technológií do praxe.

Človek mal už oddávna túžbu vytvoriť inteligentný stroj, ktorý by napodobňoval jeho myslenie a inteligenciu. Ľudstvo sa oddávna snažilo o vytvorenie umelého človeka.

Veľmi významní matematici ako aj iní vedci prispeli k tomu, aby nastal rozvoj výpočtovej techniky postavením prvého elektrónkového počítača „ENIAC" v roku 1948 a následne v roku 1956 na seminári na Univerzite v Dartmounthe v USA sa pod podporou Rockefellerovej nadácie zrodila myšlienka založenia vedného odboru „umelá inteligencia" (artificial intelligence), ktorý sa v poslednom období terminologicky označuje ako tzv. „strojová inteligencia" (machine intelligence). Seminár organizovali 4 významní predstavitelia umelej inteligencie a to Prof. Marvin Minski (za oblasť učiacich sa strojov), Prof. John McCarthy (za oblasť symbolických jazykov), Prof. Nathaniel Rochester (za neurónové siete) a Prof. Claude Shannon (za teóriu informácií).

[1] Centrum pre inteligentné technológie (CIT), skupina pre výpočtovú inteligenciu (CIG), Katedra kybernetiky a umelej inteligencie, Fakulta Elektrotechniky a Informatiky, Technická Univerzita Košice, www.ai-cit.sk, E-mail: Peter.Sincak@tuke.sk

Biologické systémy predstavujú veľmi zaujímavé technologické prostredie, kde všetky procesy fungujú na základe nejakých pravidiel, ktoré už poznáme, alebo zatiaľ ešte nie. Vo väčšine prípadov biologické systémy predstavujú deterministický a nie stochastický systém. Samotný fakt, že ľubovoľný biologický systém predstavuje veľmi zložité technologické prostredie, bol explicitne zvýraznený práve spoluzakladateľom kybernetiky Prof. Norbertom Wienerom. Táto jedinečná osobnosť svetovej vedy sa vo svojej knižnej publikácii „Kybernetika“ zaoberá problematikou regulačných a riadiacich procesov v biologických systémoch. Za základnú myšlienku možno pokladať získavanie poznatkov z biologických systémov a využívanie ich inšpiračného alebo aplikačného potenciálu.

Po seminári v roku 1956 boli ambície vedcov z oblasti strojovej inteligencie veľmi veľké, až prehnané. Myšlienka modelovania ľudského mozgu, resp. procesov myslenia človeka, bola na tú dobu veľmi predčasná a treba povedať, že aj na súčasné 21. storočie dáva veľmi problematickú ambíciu. Táto ambícia vyústila v simulácie veľmi jednoduchých procesov myslenia vo forme jednoduchých plavidlových expertných systémov, ktoré sa začali využívať aj v praxi. Myšlienka simulácie ľudského mozgu však bola neskôr zavrhnutá pre svoju zložitosť a komplexnosť problému. Teda je nutné si uvedomiť, že strojová inteligencia nemá za primárny cieľ modelovanie ľudského mozgu!!! Jej primárnym cieľom je vytvoriť prostriedky, ktoré je možné používať v rôznych technologických systémoch s cieľom zvýšenia kvality týchto technológií. Zvýšiť kvalitu technológií znamená primárnu úsporu nákladovosti technológií a tým aj finančné úspory pre prevádzkovateľa technológie. Súčasne je dôležitý aj komerčný dosah týchto technológií. Kto by nechcel kúpiť inteligentné auto, inteligentnú práčku, inteligentný vysávač alebo inteligentný prostriedok vo všeobecnosti – naviac ak to bude finančne porovnateľné s ostatnými technológiami. Čo by ste urobili Vy, ako manažér firmy, ak chcete vidieť dopredu a záleží Vám na rozvoji firmy? Minimálne by ste tieto technológie mali registrovať a poznať ich základné charakteristiky.

2 Čo sú to inteligentné technológie?

Inteligentné technológie predstavujú systémy, ktoré majú prvky strojovej inteligencie. Medzi základné vlastnosti technológií, ktoré majú strojovú inteligenciu, patria nasledovné schopnosti systémov :

- Schopnosť učenia sa z dát a získavania poznatkov
- Schopnosť ukladať poznatky
- Schopnosť využívať získané poznatky pri riešení konkrétnych situácii

Ak zavádzame nejakú technológiu do určitých podmienok, je vhodné aby sa táto technológia v daných, ako aj v nových, postupne sa meniacich podmienkach adaptovala a vedela postupne zvládať situácie, ktoré sa v nejakej forme aspoň raz počas jej činnosti vyskytli. Súčasne by už v budúcnosti mala využiť získané poznatky na riešenie podobných situácií. Táto filozofia učiacich sa systémov je veľmi zaujímavou cestou postupného vytvárania *user-friendly* technológií, ktoré sa kontinuálne učia na základe svojich skúseností, svoje znalosti uchovávajú a následne využívajú pri svojej činnosti. Tieto technológie postupom času „dozrievajú“ vzhľadom na konkrétne podmienky prevádzky. V Európe sa často takéto technológie nazývajú aj tzv. **Smart technológie**. Tieto

technológie predstavujú vyšší stupeň technológií vo všeobecnosti. Definícia inteligentných technológii nie je taká jednoznačná v rôznych komunitách. Vo väčšine prípadov ide o inovácie technológií, kde sa vkladajú do inovácie nové expertné znalosti a tým sa zvyšuje úroveň stroja. Táto úroveň stroja sa však už ďalej nemení. Teda nejde o učiace sa systémy. Veľkú časť inteligentných technológií tvoria adaptívne systémy, ktoré sa prispôsobujú rôznym podmienkam, ktoré vznikajú v rôznych situáciách, ale iba vo vybratých parametroch – teda nie komplexne.

Ak sa vrátime ku základným charakteristikám inteligentných technológií, tak je vhodné spomenúť nasledovné úvahy:

1. vlastnosť – schopnosť – učiť sa, tzv. učiace sa systémy. Je to schopnosť systému získavať informácie o jednotlivých stavoch systému, súčasne príslušných prípadných zásahoch riadiaceho systému alebo operátora v prípade, ak to riadiaci systém nezvládne a transformovať ich do znalostí. Systém vie postupne získavať znalosti dvojakého druhu:
   a) vie vytvárať usporiadané dvojice tvorené vstupmi do systému a k nim patriacimi výstupmi systému od riadiaceho systému alebo od operátora
   b) vie kategorizovať stav systému do určitých typových stavov, čo nie je také jednoduché ak je stav systému popísaný napr. niekoľkými desiatkami premenných – to znamená stav je popísaný hyper-rozmerným stavovým priestorom
2. schopnosť uchovať znalosti – takýto systém by si mal vedieť budovať bázu znalostí v ľubovoľnej forme. Ak takúto znalosť chceme využívať, musíme mať k dispozícii vhodný interpretačný prostriedok znalostí, čo nie je stále také jednoduché. Ak tento interpretačný prostriedok máme, môžeme požiadať systém o interpretáciu výsledkov. Vo všeobecnosti rozdeľujeme formu bázy znalostí do dvoch základných skupín a to tzv. štruktúrované a tzv. neštruktúrované. Štruktúrované znalosti vieme interpretovať na základe interpretačného modulu a samé osebe majú vypovedací tvar, napr. sú to symboly alebo názvy predmetov sformulované do pravidiel, apod. Naopak pri neštruktúrovanej znalosti vieme, že systém obsahuje určité znalosti, avšak tvorba interpretačného systému je veľmi zložitá až niekedy nemožná – to znamená, že znalosti v neštruktúrovanej podobe sa veľmi obtiažne interpretujú.
3. schopnosť využívať znalosti – ak sa teda systém vie učiť, následne uchovať znalosti – tak je zrejmé, že cieľom celého snaženia bude riešenie problémov v budúcnosti práve pomocou znalostí („skúseností“) získaných v minulosti. Táto vlastnosť je veľmi dôležitá a vytvára podmienky na to, aby sa používateľský („operátorský“) komfort takejto technológie postupne zlepšoval. To znamená ak vznikla v minulosti nejaká situácia v technológii a operátor ju svojim zásahom vyriešil – takýto systém by mal vedieť v budúcnosti reagovať samostatne podobne ako operátor. Je zrejmé, že takéto systémy prichádzajú s nejakou základnou bázou znalostí, ktorá sa upravuje a rozširuje. Takto stavaný systém „dozrieva“, je vysoko adaptívny a rieši väčšinu situácií v danej aplikácii.

Inteligentné technológie majú ešte množstvo ďalších vlastnosti, ale je vhodné vedieť, že uvedené vlastnosti patria medzi najdôležitejšie. Súčasne sme technologicky tak ďaleko, že pracovníci výskumu v tejto oblasti pracujú nad tvorbou štandardov v oblasti inteligentných systémov a inteligentné systémy by mali mať svoje MIQ (machine intelligent quocient). Je zrejmé, že tento MIQ bude musieť byť referenčne popisovaný na jednotný typ systémov napr. autá, vysávače, variče, vysoké pece, apod. Tento fakt je mimoriadne dôležitý pre budúcnosť aj z komerčného pohľadu a významu. Každý užívateľ resp. zákazník si bude chcieť kúpiť systém s vyšším MIQ. Technologicky je dôležité, aby si tieto jednotlivé generácie technológií odovzdávali svojej bázy znalostí, a aby bola globálna tendencia tvorby obsiahlej bázy znalosti profesne a účelovo orientovaná. Tento zámer si vyžaduje veľké úsilie, ktoré je spojené so štandardizáciou báz znalostí, čo je veľmi zložitá úloha a závislá od profesnej domény. Predstavme si, že napr. naše deti vo veku 10 rokov získajú „nejakým“ spôsobom znalosti, ktoré my ako rodičia máme napr. vo veku 40-tich rokov. To znamená, že by sa tieto znalosti replikovali a následne využívali. Je zrejmé, že u ľudí to možné nebude (čo je správne, hoci vo výskume mozgu sa takéto zámery študujú), ale u technológií sa jednotlivé bázy znalostí odovzdávať určite budú.

3 Strojová inteligencia: základné technologické pojmy

Aké sú základné terminologické pojmy, ktoré by manažér pri výbere technológii mal poznať? V oblasti strojovej inteligencie je vhodné mať základnú predstavu aspoň o nasledovných prostriedkoch strojovej inteligencie:

Klasická umelá inteligencia:

- **riešenie úloh:** riešenie úloh, zväčša založené na prehľadávaní, zahŕňa pomerne veľký problémový okruh: plánovanie, uvažovanie, rozvrhovanie, dedukcia, inferencia, dokazovanie teorém, apod. Systémy pre riešenie úloh zvyčajne obsahujú tri hlavné komponenty:
  - databáza, ktorá obsahuje dosahovaný cieľ a súčasný stav riešenia úlohy
  - množina operátorov, ktoré sa používajú na manipuláciu s databázou (napr. v šachu to môže byť množina pravidiel pre ťahanie s figúrkami)
  - stratégia riadenia: rozhoduje, ktorý operátor použiť, kedy a kde

  Pri dosahovaní stanoveného cieľa sa zvyčajne postupuje dvoma základnými spôsobmi:
  - priame uvažovanie (dátami riadené systémy): aplikácia operátorov na údajové štruktúry, ktoré reprezentujú aktuálnu situáciu (stav úlohy), kedy je úlohou pretransformovať počiatočnú situáciu na požadovanú cieľovú
  - spätné uvažovanie (cieľom riadené systémy): aplikácia operátorov na požadovanú cieľovú situáciu, kedy sa cieľ pretransformuje na jeden alebo viac podcieľov, ktoré je jednoduchšie riešiť, a ktorých riešenie je postačujúce pre vyriešenie danej úlohy. Tieto podciele sú potom ďalej redukované, až kým nie sú dosiahnuté triviálne úlohy.

Existujú samozrejme aj spôsoby využívajúce kombináciu uvedených prístupov. Výsledky tejto oblasti si našli široké uplatnenie napr. v systémoch pre výber a spracovanie informácií (information retrieval), porozumenie prirodzeného jazyka, dokazovanie matematických teorém, ale aj v oblasti robotiky a v expertných systémoch.

- **expertné systémy:** sú programy pre riešenie úloh, ktoré sú natoľko komplexné a špecifické, že ich je schopný uspokojivo riešiť iba špecialista v danom odbore – expert. Expertné systémy sú založené na myšlienke využiť znalosti prevzaté od človeka pri riešení relatívne úzko špecializovaného okruhu úloh, vyžadujúcich jeho expertné schopnosti. Pritom sa požaduje, aby riešenie, ktoré produkuje expertný systém, bolo veľmi podobné riešeniu, ku ktorému by v danom prípade dospel expert, a aby expertný systém vedel odôvodniť svoje rozhodnutie. Vo všeobecnosti možno expertný systém popísať ako systém zložený z nasledujúcich komponentov:
    - inferenčný mechanizmus: tvorí jadro expertného systému a s využitím bázy vedomostí po každej odpovedi z bázy údajov upresňuje aktuálny model konkrétneho konzultovaného prípadu a v závislosti od stavu tohto modelu kladie ďalšiu otázku. V dialógu s používateľom je iniciatíva na strane systému a sled otázok nie je pevne naprogramovaný – systém vyberá na základe doterajších odpovedí používateľa vždy najrelevantnejšiu otázku.
    - báza údajov: je tvorená odpoveďami používateľa, získanými obvykle v priebehu dialógu. V niektorých aplikáciách sa časť bázy údajov získava priamo snímaním údajov z objektu, ktorý je predmetom expertízy.
    - vysvetľovací mechanizmus: používateľ má možnosť kedykoľvek prevziať v dialógu iniciatívu a vyžiadať si vysvetlenie „uvažovania“ systému (prečo bola položená práve táto otázka, ako dospel systém k niektorému záveru, aký je momentálny stav riešenia konkrétneho problému, atď.)
    - báza vedomostí: vytvára ju znalostný inžinier na základe informácií získaných od experta a predstavuje formalizované a zakódované vedomosti ľudského špecialistu z danej problémovej oblasti.

  Expertné systémy si našli široké uplatnenie pri riešení diagnostických, klasifikačných, rozvrhovacích a plánovacích úloh, často sú rozhodovacou súčasťou zložitých riadiacich komplexov.

Výpočtová inteligencia:

- **neurónové siete:** predstavujú zjednodušený formálny model biologických neurónov a ich sietí. Činnosť neurónových sietí možno rozdeliť do dvoch základných fáz:
    - fáza učenia, kedy sa do siete ukladajú znalosti modifikáciou synaptických váh (neurónových spojení)
    - fáza života, kedy sa získané znalosti využívajú v prospech riešenia problému, pre ktorý je sieť určená

Podstatnou vlastnosťou neurónových sietí je práve učenie, ktoré možno vo všeobecnosti rozdeliť na dva základné typy:

- kontrolované učenie: učenie s učiteľom (supervised learning), kedy ku každej vstupnej učiacej (trénovacej) informácii učiteľ určí príslušnú výstupnú informáciu. Neurónová sieť sa teda snaží rozpoznávať vstupy a zatrieďovať ich tak, ako to bolo povedané učiteľom
- nekontrolované učenie: učenie bez učiteľa (unsupervised learning), kedy možno sieti v procese učenia ponúknuť iba vstupné informácie a samotná sieť určuje výstup. Prakticky ide o spracovanie vstupných informácií podľa určitých zákonitostí.

Aplikovateľnosť neurónových sietí vychádza z ich základných vlastností. Jednou z najvýznamnejších je fakt, že neurónová sieť je univerzálnym aproximátorom funkcie. Množstvo problémov je možné preformulovať ako neznáme funkcie so známymi výstupmi, príp. aj vstupmi. Neurónovú sieť potom možno chápať ako mnohoparametrický systém, ktorý sa snaží identifikovať funkciu reprezentujúcu riešený problém nastavením a adaptáciou svojich parametrov. Neurónové siete sa efektívne využívajú pri riešenie problémov v oblastiach klasifikácie, riadenia procesov, predikcie, aproximácie, spracovania signálov, apod.

- **fuzzy systémy:** sú systémy postavené na báze fuzzy logiky nad tzv. fuzzy množinami a umožňujú „presne narábať s nepresnou informáciou“. Fuzzy množina, na rozdiel od klasických množín, ktoré predstavujú presné vymedzenie, nepracuje s ostrým vymedzením. Výrazy patriť/nepatriť do množiny sú reprezentované funkciou príslušnosti, ktorá dáva fuzzy množine flexibilitu v modelovaní bežných tzv. lingvistických výrazov (napr. ak je teplota nízka, izbová, vysoká, apod., funkcia príslušnosti pre konkrétnu hodnotu premennej teploty určí, *do akej miery* je nízka, izbová, vysoká, apod.). Fuzzy systémy vo všeobecnosti pracujú s rôznymi typmi funkcií príslušnosti (trojuholníková, lichobežníková, Π-funkcia...) a používajú špeciálne operácie (rôzne typy T-noriem a T-conoriem) pre prácu s fuzzy množinami. V praxi sa bohato uplatnili kvôli možnosti práce s pravidlami typu if-then, s bázami znalostí a dát, hlavne v podobe fuzzy regulátorov a fuzzy expertných systémov.
- **systémy založené na evolúcii:** existuje množstvo rôznych typov prístupov využívajúcich evolúciu, preto sa budú v nasledujúcom texte kvôli jednoduchosti uvedené typy označovať termínom evolučné algoritmy. Evolučné algoritmy teda patria medzi prehľadávacie algoritmy, ktoré n-bodovo prehľadávajú priestor riešení za účelom nájdenia jedného, resp. viacerých riešení, využívajúc tzv. populáciu riešení (n bodov, z ktorých každý predstavuje riešenie). Jedno riešenie evolučného algoritmu je reprezentované jedným jedincom v populácii a sú na neho kladené dve základné požiadavky:

- o každý jedinec reprezentuje jedno riešenie problému pomocou genetického kódu, ktorý predstavuje parametre riešenia a ich konkrétne hodnoty
- o druhou požiadavkou je potreba určovania kvality každého riešenia, čo sa vykonáva v každom cykle riešenia úlohy, teda v každej generácii evolučného procesu. V každej generácii evolučného algoritmu sa aplikujú na jednotlivých jedincov tzv. genetické operátory (rôzne typy mutácie a kríženia; mutácia: zmena jedného parametra genetického kódu; kríženie: výmena celých častí genetického kódu), ktoré spôsobia zmenu jednotlivých riešení a zároveň zmenu ich kvality. Pomocou tzv. selekčného mechanizmu sa vyberie n najlepších riešení, ktoré postupujú do nasledujúcej generácie.

Prechod od starej množiny bodov v prehľadávanom priestore sa deje pseudonáhodne, náhodnosť nie je ponechaná „bez dozoru“. Kopírujúc evolučný princíp v prírode, náhodné zmeny (zväčša malé), ktoré vykazujú „dobrý vplyv“ na riešenie ostávajú, iné horšie zanikajú. Trvalé pôsobenie prirodzeného výberu a výberu odchýlok spôsobuje existenciu evolučného tlaku a umožňuje nájdenie najlepšieho, prípadne viacerých vyhovujúcich riešení. Systémy založené na evolúcii majú široké uplatnenie hlavne v oblasti optimalizačných úloh.

- **systémy na báze umelého života:** umelý život možno definovať ako všeobecnú metódu, ktorej podstatou je generovať z chovania jednoduchých mikroskopických spolupracujúcich prvkov také chovanie na makroskopickej úrovni, ktoré je možné interpretovať ako prejav života. Umelý život prakticky poskytuje možnosť využitia širokej škály evolučných teórií, prípadne ich kombinácií pri simulácii zložitých systémov. Virtuálne prostredie simulačných prostriedkov umelého života umožňuje na základe experimentov sledovať dlhý proces evolúcie za veľmi krátky čas. Aplikácie systémov na báze umelého života využíva na riešenie simulačných a predikčných úloh množstvo disciplín, napr. biológia, ekológia, antropológia, ekonómia, geografia, ale aj priemysel.

Na základe rýchleho vývoja v oblasti strojovej inteligencie sa javí byť dôležitým delenie systémov na základe stupňa centralizácie, resp. decentralizácie, s čím súvisí aj vznik *distribuovanej umelej inteligencie.* Táto tvorí prienik medzi distribuovanými výpočtami a umelou inteligenciou. Distribuované výpočty existujú tak dlho, ako je možné rozdeliť riešenie jedného výpočtového problému na viacero procesorov, pričom hlavnou úlohou bolo riešenie paralelizácie a synchronizácie týchto výpočtov. Distribuovaná umelá inteligencia vznikla aplikáciou distribuovaných výpočtov v prostriedkoch umelej inteligencie. Na rozdiel od distribuovaných výpočtov bola distribuovaná umelá inteligencia zameraná na riešenie problémov, koordináciu a komunikáciu oproti nízkoúrovňovej paralelizácii a syschronizácii. Rozvoj v oblasti manažmentu informácií spôsobil rozdelenie distribuovanej umelej inteligencie na paralelnú umelú inteligenciu, distribuované expertné systémy, neskôr distribuované bázy znalostí, distribuované riešenie problémov, apod. V súčasnosti sa však najčastejšie možno stretnúť s delením distribuovanej umelej inteligencie na dve triedy:

- distribuované riešenie problémov: súvisí s manažmentom informácií, dekompozíciou a distribuovaním úloh a/alebo syntézou riešení (v tomto type systémov sa kladú pomerne vysoké nároky na kompatibilitu jednotlivých riešiacich entít, ktorým sa čiastkové úlohy distribuujú)
- multiagentové systémy: súvisia s koordináciou a manažmentom správania. V súvislosti s týmto typom systémov vzniká potreba zavedenia nového pojmu: *agent*, ktorý možno vo všeobecnosti popísať ako autonómnu inteligentnú riešiacu jednotku systému. Na rozdiel od systémov distribuovaného riešenia úloh, v multiagentových systémoch nie sú kladené vysoké požiadavky na jednotlivé riešiace entity – agentov. Agenti majú tendenciu vzájomne komunikovať a kooperovať, správanie jedného agenta je zvyčajne omnoho jednoduchšie ako komplex vzájomných interakcií agentov a globálne správanie sa celého systému (skupiny interagujúcich agentov).

Centralizácia, resp. stupeň decentralizácie sa v súčasnosti javí byť jednou z kľúčových vlastností pri návrhu systému so strojovou inteligenciou. Vo všeobecnosti možno vymedziť:

- centralizované systémy: centrálne riadené systémy (tiež systémy na báze jedného agenta). Čiastkovým riešiacim jednotkám (zvyčajne špecializovaným modulom) sa distribuujú čiastkové úlohy dekomponovaného globálneho problému, ktoré sa po vyriešení odovzdávajú opäť vyšším úrovniam riadenia, kde sú syntetizované (ak je to potrebné).
- decentralizované systémy (tiež multiagentové systémy): riadenie systému je distribuované jednotlivým autonómnym agentom a je realizované na báze ich vzájomných interakcií zahŕňajúcich komunikačné a kooperatívne akty (spoločný problém je riešený kooperáciou nezávislých riešiacich jednotiek, pričom samotný priebeh riešenia spoločnej úlohy nie je riadený centrálne).

4 Potrebujeme inteligentné technológie na Slovensku?

Túto otázku môžeme interpretovať aj otázkou „Potrebujeme efektívnejšie technológie v slovenskej technickej a netechnickej praxi?“. Myslím, že čitateľ týchto riadkov má odpoveď jasnú. Aj v podmienkach Slovenskej republiky (podmienkach relatívne lacnej pracovnej sily) je potreba zvýšiť kvalitu používaných technológií. **Nie sú potrebné také inteligentné technológie, ktoré nám neprinesú zisk v úspore energie, kvalite práce a výrobkov a celkovej nákladovosti produkcie v našich podmienkach. Inteligentné technológie, ich nákladovosť a produkcia zisku závisí od konkrétnej aplikácie a ekonomických podmienok jej nasadzovania. Nenasadzujme inteligentné technológie tam, kde neprinesú ekonomický ako aj technologický prínos!**

Príklady takýchto technológií na Slovensku už máme. Ako reprezentatívne príklady možno uviesť napr. systém laminárneho chladenia vo firme US-Steel na teplej valcovni v Košiciach je realizovaný na báze neurónových sietí. Ďalším príkladom je systém kontroly kvality plechov opäť vo firme US-Steel Košice vyvinutý firmou Kybernetika s.r.o. Košice na báze neurónových sietí. Vo firme TEKO v Košiciach bol

nasadený systém na predikciu spotreby výroby tepla. Systém sa neustále učí a napomáha dispečerom pri lepšej predikcii správnej výroby tepla.

Tieto technológie sa uplatňujú aj v nezvyčajných aplikáciách. Príkladom je aj napr. hrajúca fontána v Košiciach, kde do tejto technológie boli zabudované prvky tzv. fuzzy logiky.

Použitie systémov so strojovou inteligenciou nie je vo svete žiadnou novinkou. Tieto technológie sú úspešne nasadzované do praxe firmami ako napr. Hitachi, Matsushita, Mitsubishi, Sharp, Rockwell Corp. Nissan, Honda, Subaru, Fuji, Canon, Isuzu, Nissan, Fujitec, Sanyo, Minolta, NASA, Goldstar, Hitachi, Samsung, Sony, Toshiba, Sharp a mnohými ďalšími.

Kognitívna sémantika komplexných kategórií založená na rozlišovacích kritériách

Martin Takáč[1]

Abstrakt. V článku prezentujeme sémantiku vhodnú na reprezentáciu rôznych typov konceptov (objektov, vlastností, vzťahov, zmien, udalostí a situácií) v dynamických a otvorených prostrediach pre kognitívne modely vývoja a osvojovania jazyka. Sémantika je založená na perceptuálne ukotvených štruktúrach – rozlišovacích kritériách, ktoré sú osvojiteľné na základe interakcií s prostredím a inými agentmi.

1 Úvod

Osvojovanie si jazyka sa tradične chápe ako problém získania správnych asociácií medzi externe pozorovateľnými formami (slovami, vetami, gestami) a internými významami [23]. Významy, ako mentálne koncepty, sa buď považujú za vrodené [10], získané interakciou so svetom [5] alebo vytvorené čisto vplyvom samotného jazyka [40]. Ak by sme prijali extrémny názor, že všetky koncepty sú vrodené, osvojovanie si jazyka by spočívalo iba v naučení sa nových slov (nálepiek) pre existujúce významy. Ak by sme prijali opačný extrém, bez jazyka by nemohlo existovať žiadne myslenie.

Empirické výskumy napovedajú, že deti majú rôzne poznatky o svete omnoho skôr, ako začnú hovoriť [15], osvojujú si rôzne zručnosti a senzomotorické schémy [28] a rozlišujú medzi rôznymi objektmi a situáciami [29]. Na evolučnej scéne sa jazyk objavuje pomerne neskoro, napriek tomu sú mnohé živočíšne druhy schopné kategorizovať (napr. identifikovať objekty vhodné na zjedenie alebo vyhodnotiť situáciu ako nebezpečnú). Slovo koncept, resp. kategória[2] budeme odteraz používať na označenie abstrakcie schopnosti rozlišovať medzi objektmi, situáciami, stavmi, atď. [31] (teda ak napr. zviera dokáže isté objekty identifikovať ako jedlo, povieme, že disponuje kategóriou jedla). Z takejto definície vyplýva, že (aspoň niektoré) koncepty môžu existovať aj nezávisle od jazyka. V súvislosti s jazykom však hrajú dôležitú úlohu, pretože slúžia ako významy jazykových výrazov, a poskytujú tak potrebné ukotvenie symbolov [16]. V tomto zmysle sú koncepty stelesnené neverbálne štruktúry späté s perceptuálnou skúsenosťou a fyzicky korešpondujú s aktiváciami neurálnych okruhov korelujúcich s vnímaním, konaním, predstavovaním si alebo rozprávaním a počúvaním o obsahu, ktorý reprezentujú [3]. Kognitívna sémantika sa zaoberá skúmaním a formalizáciou takýchto štruktúr [11, 23, 35].

Podľa nášho názoru je dôležité, aby sémantika spolu s návrhom reprezentácie konceptuálnych štruktúr opisovala aj mechanizmy ich osvojenia. Takto poňatá sémantika môže nájsť uplatnenie nielen v kognitívnych modeloch, ale aj v industriálnych aplikáciách. Moderné technológie založené na autonómnych agentoch, ktoré navzájom vyjednávajú, vymieňajú si dáta a koordinujú svoje aktivity v otvorených a dynamických

[1] Katedra aplikovanej informatiky FMFI UK, Mlynská dolina, 842 48 Bratislava,
E-mail: takac@ii.fmph.uniba.sk

[2] Výrazy koncept a kategória budeme v texte voľne zamieňať, používame ich v rovnakom význame.

prostrediach, kde žiadna preddefinovaná ontológia či jazyk nemôže postihnúť vopred všetky potenciálne komunikačné témy, si vyžadujú agenty schopné vytvoriť si a kontinuálne aktualizovať ontológie relevantné prostrediu a vzájomným interakciám a vytvoriť si vlastný komunikačný systém.

Väčšina multiagentových modelov emergencie jazyka (prehľad v [22, 30]) však neposkytuje dostatočné prostriedky na reprezentáciu dynamických konceptov ako sú procesy, udalosti, akcie a situácie, ktoré sú dôležité pre sémantiku slovies. V kognitívnej sémantike konceptuálnych priestorov [11] sú koncepty reprezentované ako oblasti v mnohorozmernom geometrickom priestore. Riešenie sémantiky slovies pridaním časovej dimenzie je však problematické [12, 31].

V tomto článku prezentujeme novú sémantiku, ktorá sa snaží prekonať vyššie opísané nedostatky. Sémantika je založená na perceptuálne ukotvených konceptuálnych štruktúrach – rozlišovacích kritériách, ktoré sú jednotiacim formálnym rámcom pre reprezentácie objektov, vlastností, vzťahov, zmien, udalostí a situácií. Významy slovies sú reprezentované ako medzikategoriálne asociácie kritérií. V článku definujeme a analyzujeme vzájomné vzťahy perceptuálnej, reprezentačnej, jazykovej a pragmatickej úrovne ako aj dôsledky spriahnutia procesov formovania kritérií a osvojovania jazyka.

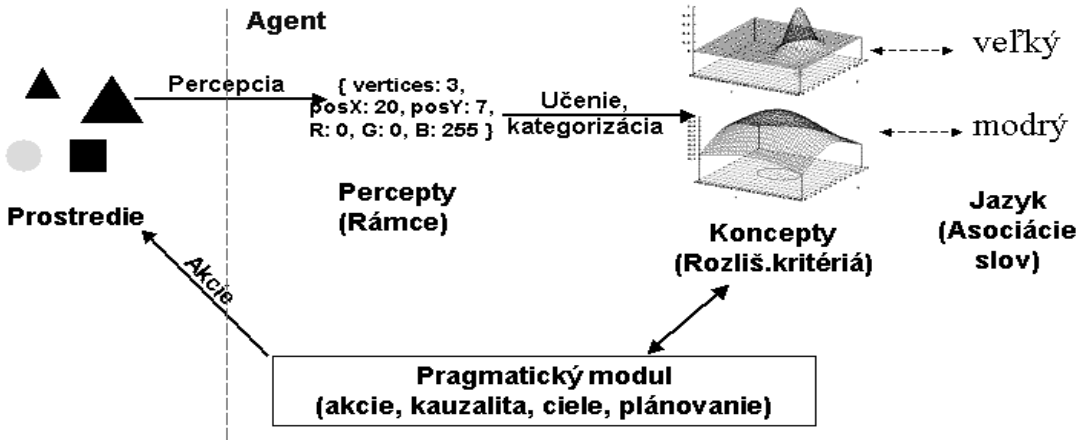
2 Od percepcie ku kategóriám a správaniu v prostredí

Akýkoľvek agent – améba, webovský robot alebo človek, reagujúci na isté stavy svojho prostredia istým spôsobom, potrebuje rozlišovať medzi stavmi prostredia tie, ktoré sú relevantné jeho cieľom. Táto elementárna schopnosť považovať niektoré stavy prostredia za identické (vzhľadom na nejaké kritériá) a odlišovať ich od iných stavov, je jadrom každej kategorizácie.

V najjednoduchšom prípade čisto reaktívnych agentov chýba akákoľvek vnútorná reprezentácia a celé správanie je realizované formou priamych spojov „stimul-reakcia“ asociujúcich perceptuálne vstupy s behaviorálnym výstupom. V tomto článku sa chceme venovať zložitejším kognitívnym agentom vykonávajúcim úlohy vyžadujúce flexibilitu, učenie, vzájomnú kooperáciu a komunikáciu v dynamických prostrediach. V ďalšom texte použijeme štyri úrovne opisu (obr. 1):

- **Perceptuálna úroveň.** Táto úroveň tvorí rozhranie medzi externým prostredím agenta a vyššími úrovňami. V stelesnených agentoch zodpovedá signálom zo senzorov agenta predspracovaným nízkoúrovňovými perceptuálnymi rutinami. V softvérových agentoch zodpovedá vstupným dátam, na ktorých agent operuje, opísaným formou, ktorá je spracovateľná na reprezentačnej úrovni.
- **Reprezentačná úroveň.** Toto je úroveň kategórií, resp. konceptov. Každý koncept je reprezentovaný rozlišovacím kritériom – funkciou, ktorá vracia hodnotu pravdepodobnosti, že jej vstup je inštanciou reprezentovaného konceptu.
- **Jazyková úroveň.** Rozlišovacie kritériá agenta sú súkromné a nie sú priamo prenositeľné ostatným agentom. Agenty komunikujú pomocou konvenčne ustanovených signálov jazykovej úrovne. Význammi jazykových signálov sú perceptuálne ukotvené kritériá z reprezentačnej úrovne. Agenty môžu úspešne komunikovať, iba ak sú ich súkromné významy dostatočne podobné. To môže nastať, pokiaľ agenty používajú podobné mechanizmy formovania konceptov na základe podobných skúseností v zdieľanom prostredí.

- **Pragmatická úroveň.** Na tejto úrovni agenty konajú v prostredí, plánujú a dosahujú ciele. Potrebné kauzálne znalosti o vlastných akciách a ich dôsledkoch sú reprezentované medzikategoriálnymi asociáciami kritérií, ciele sú reprezentované kritériami želaných situácií a plány sekvenciami akcií, ktoré by mali viesť z aktuálnej situácie do cieľa [31].



Obrázok 1. Kognitívna architektúra agenta zahŕňa percepčnú, reprezentačnú, jazykovú a pragmatickú vrstvu.

3 Perceptuálna úroveň

Budeme predpokladať, že dynamické prostredie agentov sa mení v diskrétnych krokoch. V každom kroku každý agent vníma svoje okolie, aktualizuje si vnútornú reprezentáciu a môže komunikovať alebo vykonávať akcie v prostredí podľa konkrétnej aplikácie. Perceptuálny vstup agenta v danom čase – *scénu* – opíšeme množinou rámcov zložených z párov atribút-hodnota, napr.

```
{vrcholy:3; velkost:5; posX:12; posY:18}.
```

Rámce reprezentujú napr. fyzické objekty v okolí agenta, prichádzajúce dáta, alebo agentove „proprioceptívne“ vnemy (hodnoty vnútorných stavových premenných, parametre vykonávaných akcií, pozíciu ramena, atď.). Formálne je perceptuálny rámec f charakterizovaný množinou atribútov A_f a selektorom atribútov $h_f : A_f \rightarrow \mathbf{R}$ ($\mathbf{R}$ označuje množinu reálnych čísel). V texte budeme namiesto $h_f(a)$ používať zaužívanejšiu notáciu $f.a$.

Používanie rámcov alebo iných arbitrárnych amodálnych symbolov na sémantickú reprezentáciu kritizoval Barsalou [3]. V našom modeli nepoužívame rámce ako reprezentáciu; sú len výpočtovým opisom štruktúr, ktoré vznikajú ako výsledok nízkoúrovňového predspracovania senzorického vstupu agenta. Môžu opisovať napr. pole intenzít zo sietnice alebo web-kamery alebo segmentáciu scény na objekty a ich perceptuálne charakteristiky. Ak dva agenty vnímajú rovnakú scénu, dostávajú identické množiny vstupných rámcov, ďalej ich však môžu spracovávať, kategorizovať a reprezentovať rôzne.

4 Reprezentačná úroveň

Agent sa postupne učí rozlišovať vlastnosti prostredia a združuje do kategórií entity, ktoré sú z nejakého hľadiska podobné. Každá kategória je reprezentovaná tzv. rozlišovacím kritériom – aktivačnou funkciou, ktorej hodnota vyjadruje stupeň príslušnosti vstupu funkcie k reprezentovanej kategórii. Vstupom funkcie môže byť rámec objektu (u kritérií objektov a vlastností), viacerých objektov (u kritérií vzťahu), rámce toho istého objektu v rôznych časoch (u kritérií zmeny) aj výsledky iných rozlišovacích kritérií (u kompozičných kritérií situácií a udalostí). Na začiatku agent nemá žiadne kritériá – tie sú budované na základe príkladov extrakciou ich spoločných štatistických vlastností.

Extrakciu štatistických vlastností z príkladov a ich použitie na kategorizáciu nových príkladov zabezpečujú tzv. *lokálne reagujúce detektory*, ktoré sú jadrom každého rozlišovacieho kritéria. Lokálne reagujúci detektor dostáva ako vstupný argument jediný rámec a vracia hodnotu z uzavretého intervalu [0, 1] vyjadrujúcu, do akej miery je rámec inštanciou reprezentovaného konceptu (1 znamená najlepší príklad – prototyp).

4.1 Geometrický pohľad na kategórie

Lokálne detektory majú intuitívnu geometrickú interpretáciu založenú na konceptuálnych priestoroch [11]. Konceptuálny priestor je geometrický priestor s dimenziami zodpovedajúcimi atribútom reprezentovaných entít. Keďže nie všetky atribúty sú relevantné pre každú reprezentovanú entitu, dimenzie sú organizované do podpriestorov – domén. Konkrétny objekt je reprezentovaný ako bod v priestore zloženom z jednej alebo viacerých domén a podobnosť medzi dvoma objektmi je nepriamo úmerná vzdialenosti ich bodov (aby bolo možné vypočítať vzdialenosť, body musia zdieľať podpriestor so spoločnou metrikou). Reprezentácia prirodzených kategórií je založená na predpoklade konvexnosti: ak dva body reprezentujú objekty, ktoré sú dobrým príkladom nejakej kategórie, aj všetky body ležiace medzi nimi by mali byť dobrým príkladom tejto kategórie [11]. Prirodzené kategórie sú teda reprezentované konvexnými oblasťami v konceptuálnom priestore. Aktivita lokálne reagujúceho detektora bude tým vyššia (bližšia k jednej), čím je jeho vstup lepším príkladom reprezentovanej kategórie. Lokálne reagujúci detektor môže reprezentovať kategórie s neostrými (fuzzy) hranicami, ale z praktických dôvodov môžeme stanoviť rozhodovací prah detektora θ pre vymedzenie hranice kategórie. Oblasť vstupného priestoru D, pre ktorú detektor r vracia nadprahovú aktivitu, nazveme jeho *receptívnym poľom* $\Psi_\theta(r) = \{ x \in D \mid r(x) > \theta \}$. Lokálne reagujúce detektory majú vysokú neurálnu a biologickú plauzibilitu [2,17,26].

4.2 Indukcia lokálne reagujúceho detektora z príkladov

V tejto časti opíšeme rôzne verzie mechanizmu vytvorenia detektora na základe extrakcie štatistických vlastností súboru vstupných príkladov. Čitateľ, ktorý sa nezaujíma o matematické detaily, alebo je preňho použitý matematický aparát priťažký, môže preskočiť na časť 4.3.

Predpokladajme, že agent má odvodiť reprezentáciu neznámeho konceptu z príkladov, teda vytvoriť lokálne reagujúci detektor zo sekvencie rámcov $\{f^{(1)}, ..., f^{(N)}\}$.[3] Každý rámec $f^{(i)}$ zodpovedá bodu $(f^{(i)}.a_1, ..., f^{(i)}.a_k^{(i)})$ v konceptuálnom priestore s dimenziami definovanými atribútmi $a_j \in A_f^{(i)}$, kde $k^{(i)} = |A_f^{(i)}|$ je ich počet. Indukciu založíme na štatistických vlastnostiach hodnôt atribútov spoločných pre všetky rámce, teda najprv urobíme projekciu všetkých rámcov do spoločného podpriestoru $\mathcal{A}$ definovaného prienikom množín atribútov $\bigcap_{i=1}^{N} A_{f^{(i)}}$. Atribúty, ktoré sa nevyskytujú vo všetkých rámcoch, budeme považovať za irelevantné pre príslušnosť ku kategórii. Preto môžeme teraz príklady kategórie reprezentovať ako množinu vektorov $\vec{x}^{(i)} = (x_1^{(i)},...,x_n^{(i)})$ pre $i = \overline{1,N}$ v spoločnom priestore $\mathcal{A}$ s dimenzionalitou $n = \left| \bigcap_{i=1}^{N} A_{f^{(i)}} \right|$.

Za najlepší príklad – prototyp kategórie budeme považovať geometrické centrum, resp. priemer všetkých príkladov

$$\vec{p} = \frac{1}{N}\sum_{i=1}^{N} \vec{x}^{(i)} .$$

Lokálne reagujúci detektor r_p založený na prototype vráti aktiváciu exponenciálne klesajúcu s vzdialenosťou od prototypu

$$r_{\vec{p}}(\vec{x}) = exp(-k \cdot d(\vec{p},\vec{x})), \tag{1}$$

kde k je kladná konštanta, $d(\vec{p},\vec{x})$ je vhodne zvolená metrika a $\vec{x}$ je projekcia vstupného rámca do priestoru $\mathcal{A}$ (ak sa rámec nedá projektovať, pretože v ňom chýbajú niektoré atribúty z $\mathcal{A}$, detektor vráti nulu).

Tvar receptívneho poľa $\Psi_\theta(r) = \{ \vec{x} \in \mathcal{A} \mid r_{\vec{p}}(\vec{x}) > \theta \}$ v priestore $\mathcal{A}$ závisí od použitej metriky. V prípade Euklidovskej metriky d_{L_2} je to n-rozmerná hyperguľa so stredom $\vec{p}$ a polomerom θ.

Pri použití jediného rozhodovacieho prahu θ by mali receptívne polia všetkých detektorov rovnaký polomer bez ohľadu na distribúciu príkladov, z ktorých boli odvodené. To nie je žiaduce, ak potrebujeme reprezentovať kategórie s rôznou mierou všeobecnosti. Najjednoduchším vylepšením je použiť funkciu vzdialenosti váhovanú prevrátenou hodnotou rozptylu σ^2 hodnôt všetkých atribútov v množine príkladov

$$d_{L_2,\sigma}(\vec{p},\vec{x}) = \sqrt{\sum_{i=1}^{n} \frac{1}{\sigma^2}(x_i - p_i)^2} = \frac{1}{\sigma} \cdot d_{L_2}(\vec{p},\vec{x})$$

Teraz majú kategórie s väčším rozptylom väčšie receptívne pole. Ak povolíme nekonečne veľké váhy a dodefinujeme $\infty \cdot 0 = 0$, receptívne pole kategórie s nulovým rozptylom bude jediný bod (prototyp $\vec{p}$). Takéto kategórie reprezentujú indivíduá.

[3] V praxi je to zložitejšie, agent v priebehu času vníma zmiešanú sekvenciu inštancií rôznych konceptov. Dokonca jeden perceptuálny rámec môže byť dobrou inštanciou viacerých konceptov, napr. jedno konkrétne jablko môže byť príkladom konceptov označených slovami „červený“, „okrúhly“, „malý“, „jablko“. Problémom určenia, ktorý percept je inštanciou ktorého konceptu, sa zaoberáme v časti 5.

Prirodzeným rozšírením predchádzajúceho prípadu je zaznamenávať rozptyly zvlášť pre každú dimenziu a použiť váhovanú euklidovskú vzdialenosť $d_{L_2,\vec{w}}$ s vektorom váh $\vec{w} = \left(\frac{1}{\sigma_1^2},...,\frac{1}{\sigma_n^2}\right)$, teda $d_{L_2,\vec{w}}(\vec{p},\vec{x}) = \sqrt{\sum_{i=1}^{n}\frac{(x_i - p_i)^2}{\sigma_i^2}}$.

Receptívne pole je v tomto prípade *n*-rozmerná hyperelipsa (elipsoid) so stredom $\vec{p}$ a dĺžkami osí úmernými σ_i, pričom osi sú rovnobežné so súradnicovými osami priestoru $\mathcal{A}$. Pre príslušnosť ku kategórii majú teda väčšiu dôležitosť atribúty, v hodnotách ktorých sa vstupné príklady veľmi nelíšia, ako atribúty s veľkým rozptylom. Ak sa v hodnote niektorého atribútu zhodujú všetky príklady (teda má nulový rozptyl), táto hodnota sa stane pre príslušnosť ku kategórii povinnou (akákoľvek iná dá nekonečnú vzdialenosť, a teda nulovú aktivitu detektora). Receptívne pole bude degenerovaný elipsoid s príslušnou osou nulovej dĺžky.

Váhované dimenzie sa používajú aj v pôvodnej sémantike konceptuálnych priestorov [11] na vyjadrenie kontextovej dôležitosti jednotlivých dimenzií. Rozdiel je v tom, že v našom prístupe sú váhy lokálne pre každý detektor, čím sa posudzovanie podobnosti stáva asymetrickým, teda typický príklad jednej kategórie môže byť lepšou inštanciou druhej kategórie, ako naopak[4] (vyjadrené matematicky, pre dva detektory $r_{\vec{p}_1}$ a $r_{\vec{p}_2}$ sa môžu hodnoty $r_{\vec{p}_1}(\vec{p}_2)$ a $r_{\vec{p}_2}(\vec{p}_1)$ líšiť).

Detektory zohľadňujúce dôležitosť jednotlivých atribútov na základe rozptylov sa nedokážu naučiť korelácie medzi atribútmi, zatiaľ čo ľudia túto úlohu úspešne zvládajú [27]. Napríklad pre odvodenie konceptu pre štvorec je nutné naučiť sa nielen povinnú hodnotu atribútu `pocet_vrcholov:4`, ale aj to, že atribúty `strana1` a `strana2` musia byť rovnaké. Toto dokáže zvládnuť detektor používajúci Mahalanobisovu metriku

$$d^2_{\Sigma^{-1}}(\vec{p},\vec{x}) = (\vec{x} - \vec{p})^{\mathrm{T}}\Sigma^{-1}(\vec{x} - \vec{p}),$$

kde $\vec{x}$ a $\vec{p}$ sú stĺpcové vektory a Σ^{-1} je inverzná matica ku kovariančnej matici Σ množiny príkladov konceptu.

$$\Sigma = \left(\sigma_{ij}\right)_{i,j=\overline{1,n}}, \text{ kde } \sigma_{ij} = \frac{1}{N-1}\sum_{k=1}^{N}(x_i^{(k)} - p_i)(x_j^{(k)} - p_j).$$

Keďže príklady kategórie nie sú obvykle k dispozícii všetky naraz, ale postupne, detektor bude používať na inkrementálny výpočet priemeru a kovariančnej matice iteratívne formuly. Nech *N*–1 je počet doteraz videných príkladov a $\vec{x}^{(N)}$ je nový (*N*-tý) príklad. Potom:

- pre $N = 1$ je $\vec{p}^{(1)} = \vec{x}^{(1)}$ a $\Sigma^{(1)} = (0)_{n\times n}$ alebo $\sigma_0^2 I_n$, kde σ_0^2 je nejaký počiatočný odhad rozptylu,
- pre $N > 1$ použijeme formuly

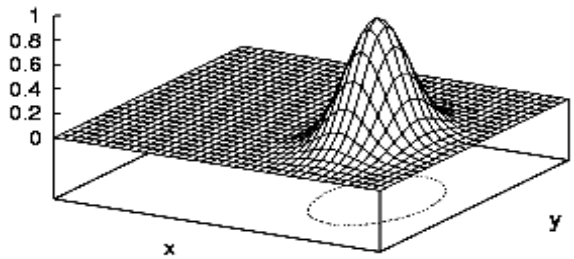
$$\vec{p}^{(N)} = \frac{N-1}{N}\vec{p}^{(N-1)} + \frac{1}{N}\vec{x}^{(N)}$$

$$\Sigma^{(N)} = \frac{N-2}{N-1}\Sigma^{(N-1)} + \frac{N}{(N-1)^2}\left(\vec{x}^{(N)} - \vec{p}^{(N)}\right)\left(\vec{x}^{(N)} - \vec{p}^{(N)}\right)^{\mathrm{T}}.$$

[4] Ľudia tiež posudzujú podobnosti asymetricky, napr. tvrdia, že Tel Aviv sa podobá na New York viac, ako sa podobá New York na Tel Aviv [38].

Množina atribútov určujúca priestor $\mathcal{A}$ sa tiež aktualizuje iteratívne prienikom s množinou atribútov nového príkladu. Ak sa takto odstránia niektoré dimenzie z $\mathcal{A}$, príslušný riadok a stĺpec sa odstráni z kovariančnej matice, resp. z prototypového vektora.

Ak v rovnici (1) položíme k = 0.5, dostaneme detektor, ktorého aktivačná funkcia má tvar mnohorozmerného (nenormalizovaného) Gausiánu (obr. 2).



Obrázok 2. Lokálne reagujúci detektor v 2-rozmernom vstupnom priestore, s aktivačnou funkciou v tvare mnohorozmerného Gausiánu. Receptívne pole detektora pre prah θ = 0.1 je zobrazené v rovine (x, y).

Keďže Σ je štvorcová symetrická kladne semidefinitná matica, možno ju rozložiť na súčin matíc

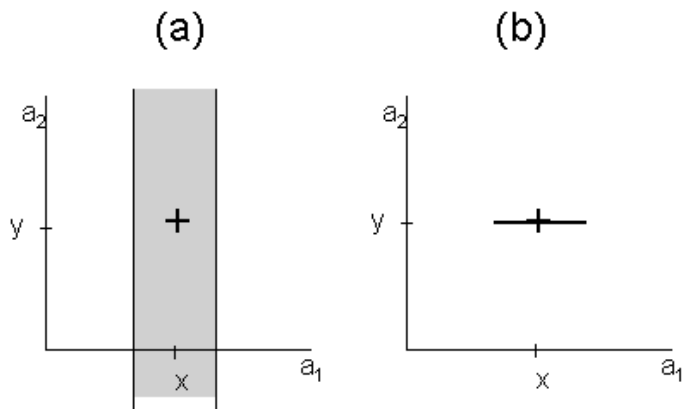
$$\Sigma = UDU^T, \tag{2}$$

kde U je ortonormálna matica rotácie a D = diag(λ_1, ..., λ_n) je diagonálna matica vlastných hodnôt matice Σ, pričom $\lambda_1 \geq \ldots \geq \lambda_n$ [20]. Potom vzdialenosť $d^2_{\Sigma^{-1}}(\vec{p},\vec{x}) = (\vec{x}-\vec{p})^T\Sigma^{-1}(\vec{x}-\vec{p}) = (U(\vec{x}-\vec{p}))^T D^{-1}(U(\vec{x}-\vec{p}))$ je vlastne vzdialenosťou medzi vektormi $\vec{x}$ a $\vec{p}$ v priestore zrotovanom podľa U, kde rozdiely v jednotlivých dimenziách sú škálované podľa D^{-1}. Keďže rotácia nemení tvar ani veľkosť receptívneho poľa, vidíme, že receptívnym poľom bude hyperelipsa, ktorej osi majú dĺžky úmerné λ_i, ale nemusia byť rovnobežné so súradnicovými osami priestoru $\mathcal{A}$. Metriky d_{L_2}, $d_{L_2,\sigma}$, resp. $d_{L_2,\vec{w}}$ sú špeciálnymi prípadmi metriky $d_{\Sigma^{-1}}$ s kovariančnými maticami I_n, $\sigma^2 I_n$, resp. diag(λ_1, ..., λ_n).

Ak sa v množine príkladov vyskytuje atribút s nulovým rozptylom, matica Σ bude singulárna, teda neinvertovateľná. V praxi sa namiesto inverznej matice Σ^{-1} používa *Moore-Penrosova pseudoinverzná matica* Σ^+, ktorej výpočet je založený na rozložení na singulárne hodnoty (*Singular Value Decomposition* – SVD) matice Σ, ktorá má podobu (2). V prípade singulárnej matice Σ je D = diag(λ_1, ... , λ_k, 0, ... ,0) pre nejaké $k < n$. Potom $\Sigma^+ = UD^+U^T$, kde D^+ = diag($1/\lambda_1$, ... , $1/\lambda_k$, 0, ... ,0) .

Detektor používajúci Σ^+ bude atribúty konštantné v celej vzorke ignorovať, namiesto aby ich považoval za povinné: ich váha bude nulová a zodpovedajúce osi degenerovanej hyperelipsy budú mať nekonečnú dĺžku (obr. 3). To je proti filozofii zachytávania regularít vo vstupnej vzorke. Napr. konštitutívnou vlastnosťou kategórie *trojuholník* je „mať 3 vrcholy“. Ak však budú mať všetky príklady vo vstupnej vzorke 3 vrcholy, kovariančná matica bude singulárna a atribút *početVrcholov* sa bude ignorovať, lebo v D^+ dostane nulovú váhu. Preto budeme namiesto D^+ používať v SVD forme matice Σ štandardnú inverznú maticu D^{-1} = diag($1/\lambda_1$, ... , $1/\lambda_k$, ∞, ... ,∞) s povolením nekonečných prvkov $1/\lambda_i = \infty$ pre $\lambda_i = 0$. Výsledok funkcie vzdialenosti bude konečný, len

ak bude mať otočený vektor $U(\vec{x} - \vec{p})$ na osiach zodpovedajúcich nekonečným prvkom D^{-1} nulové súradnice.



Obrázok 3. Príklad kategórie s prototypom (x, y) a nulovým rozptylom v atribúte a_2. Detektor používajúci pseudoinverznú maticu bude považovať atribút a_2 za nepodstatný (a), zatiaľ čo detektor používajúci inverznú maticu s nekonečnými prvkami bude považovať hodnotu $a_2 = y$ za povinnú (b).

SVD dekompozícia je užitočná aj pre zovšeobecnenie výberom konštitutívnych čŕt kategórie. Napríklad pri odvodzovaní pojmu „okrúhly“ môže pozorovateľ zistiť, že príklady okrúhlych vecí majú v porovnaní s tvarom veľkú variabilitu farieb, veľkostí a iných vlastností, na základe čoho ich môže považovať za nepodstatné. V našom modeli tento typ zovšeobecnenia zodpovedá nájdeniu tých prvkov matice D^{-1}, ktoré sú v porovnaní s ostatnými veľmi malé, a nahradeniu ich nulami. Pretože $1/\lambda_1 \leq ... \leq 1/\lambda_n$, môžeme nájsť najväčšie L také, že

$$\frac{\sum_{i=1}^{L} \frac{1}{\lambda_i}}{\sum_{i=1}^{n} \frac{1}{\lambda_i}} < b, \tag{3}$$

kde b je nejaký percentuálny prah, napr. $b = 10\%$. Nahradením prvých L prvkov D^{-1} nulami zanedbáme komponenty prispievajúce k celkovej vzdialenosti len veľmi málo. Toto je komplementárny proces k metóde *Principal Component Analysis* (PCA), v ktorej sa dátový priestor redukuje na komponenty s najväčšími rozptylmi [18].

Z rovnice (3) vyplýva, že ak matica D^{-1} obsahuje nekonečné prvky (zodpovedajúce úplne invariantným vlastnostiam vstupnej vzorky), všetky konečné sa zanedbajú. To je správne pre väčšinu konceptov, avšak nekonečné prvky môžu byť len umelou chybou – artefaktom spôsobeným niektorými konštantnými, ale úplne irelevantnými atribútmi (napr. ak hľadáme medzi predmetmi položenými na stole invariantné vlastnosti všetkých jabĺk a zaradíme medzi ne aj to, že všetky sú na stole). Odlíšenie figúry od pozadia je spoločným problémom všetkých indukčných algoritmov, ktoré sa učia iba z pozitívnych príkladov [14]. Vlastnosť zdieľaná všetkými pozitívnymi príkladmi by sa mala považovať za irelevantnú, ak ju zdieľajú aj negatívne príklady (ignorovaniu spoločných vlastností zodpovedá už zmienené použitie pseudoinverznej matice s D^{+}). V našom modeli sa informácia o negatívnych príkladoch nevyužíva.

4.3 Znamienkové lokálne detektory

V každodennom živote ľudia často abstrahujú od numerických hodnôt a rovníc a používajú jednoduchší kvalitatívny kalkul založený na ordinálnych vzťahoch a invariantných znamienkových štruktúrach [32]. Znamienková štruktúra atribútov je konštitutívna aj pre mnohé vzťahové koncepty, napr. relácia *vacsiAko*(f_1, f_2) je založená na vlastnosti $f_1.velkost > f_2.velkost$, resp. sgn($f_1.velkost - f_2.velkost$) = +1.

Na rozlišovanie znamienkovej štruktúry atribútov môžeme zaviesť znamienkový lokálne reagujúci detektor. Podpriestor $\mathcal{A}$ bude definovaný množinou tých atribútov, ktoré sa vo všetkých príkladoch vstupnej vzorky vyskytujú s rovnakým znamienkom a projekcia do podpriestoru $\mathcal{A}$ bude zložená s operátorom sgn. Spoločná znamienková štruktúra príkladov vstupnej vzorky bude zaznamenaná v prototype $\vec{p} = (p_1, ..., p_n)$, kde $p_j = sgn\left(x_j^{(i)}\right) \quad \forall i = \overline{1, N}$. Znamienková štruktúra sa zaznamená iba raz, po zhliadnutí prvého príkladu. Neskoršie aktualizácie kritéria iba odstraňujú z $\mathcal{A}$ atribúty, ktoré majú v nových príkladoch iné znamienko ako to, čo je zaznamenané v prototype. Detektor aplikovaný na nejaký vstup potom reaguje binárnou aktivitou: ak sa projekcia vstupu do $\mathcal{A}$ zhoduje s prototypom vo všetkých zložkách, vráti hodnotu 1, inak vráti hodnotu 0.[5]

4.4 Implementácia rozlišovacích kritérií pomocou detektorov

Rozlišovacie kritériá individuálnych objektov, napr. *Bratislava*(*f*), vlastností objektov, napr. *modrý*(*f*), *veľký*(*f*) alebo *zadĺžený*(*f*), a tried objektov, napr. *strom*(*f*), *dieťa*(*f*), alebo *počítač*(*f*),[6] môžeme priamo implementovať lokálne reagujúcim detektorom, pretože ich vstupným argumentom je jediný rámec. Z formálneho hľadiska nie je medzi kritériami indivíduí, vlastností a tried žiadny rozdiel.

Rozlišovacie kritériá, ktoré majú viac vstupných argumentov (napr. vzťahy a zmeny) sa dajú tiež redukovať na lokálne reagujúce detektory transformáciou ich vstupu na jeden rámec. Binárne vzťahy medzi objektmi, napr. *väčšíAko*(f_1, f_2) alebo *blízko*(f_1, f_2), sa dajú implementovať lokálne reagujúcim detektorom, ktorého vstupom bude rámec $f = \Delta(f_1, f_2)$ rozdielov hodnôt atribútov prítomných v oboch rámcoch: $A_f = A_{f_1} \cap A_{f_2}$ a $f.a = f_1.a - f_2.a$ pre každé $a \in A_f$.

Na reprezentovanie vzťahov založených na magnitúde rozdielov (napr. *blízko*, *omnohoViac*, *takmerRovnaký*), je vhodné použiť detektor založený na kvantitatívnych štatistických vlastnostiach vstupného súboru, napr. priemere a kovarianciách. Na reprezentáciu kvalitatívnych vzťahov založených na znamienku rozdielov atribútov, resp. usporiadaní ich hodnôt je vhodné použiť znamienkový lokálny detektor (pozri časť 4.3) aplikovaný na vstup $\Delta(f_1, f_2)$.

---

[5] Znamienkový detektor možno chápať ako špeciálny prípad detektora s metrikou $d_{L_2,\vec{w}}$, v ktorom sú hodnoty (znamienka) všetkých atribútov povinné (keďže majú vo vstupnom súbore nulový rozptyl).

[6] Použité identifikátory rozlišovacích kritérií majú čitateľovi napovedať, aký koncept je reprezentovaný kritériom, nejedná sa však o zápis jazykového výrazu. Jazykové výrazy budeme v texte od označenia rozlišovacích kritérií odlišovať typom písma a úvodzovkami, napr. „veľký“ je slovo (a jeho významom môže byť rozlišovacie kritérium *veľký*).

Kritériá na rozlišovanie situácií zahŕňajúcich komplexnejšie vzťahy medzi viac ako dvoma objektmi alebo vlastnosti celej scény sa dajú vybudovať ako hierarchické siete lokálne reagujúcich detektorov. Kritériá na najnižšej úrovni hierarchie rozlišujú objekty, ktoré majú byť prítomné na scéne a ich požadované vzájomné vzťahy. Vektor výstupných aktivít týchto kritérií slúži ako vstup pre hierarchicky vyššie kritériá, ktoré môžu rozlišovať požadované hodnoty aktivít jednotlivých elementárnejších kritérií s rôznou dôležitosťou, detekovať vzájomné korelácie aktivít a podobne.

4.5 Reprezentácia dynamiky prostredia: kritériá zmeny

Pre agenty operujúce v dynamických prostrediach je veľmi dôležitá schopnosť vnímať a reprezentovať zmeny. V spojite sa meniacom svete dokážu už štvormesačné deti sledovať objekty pohľadom a osvojiť si princíp *kontinuity objektov* [21], ktorý je nutnou podmienkou pre vnímanie zmien vlastností objektov.

Keďže v našom modeli sa čas mení v diskrétnych krokoch, musíme zabezpečiť, aby agent nevnímal scény v jednotlivých krokoch ako nezávislé, ale ako množiny s korešpondenciami medzi perceptuálnymi rámcami toho istého objektu v rôznych časoch. Z tohto dôvodu predefinujeme *scénu* – perceptuálny vstup agenta v čase *t* ako množinu dvojíc perceptuálnych rámcov a ich podôb v predchádzajúcom časovom kroku:

$$S_t = \left\{ \left(f_1^{(t)}, f_1^{(t-1)}\right), \ldots, \left(f_k^{(t)}, f_k^{(t-1)}\right) \right\}.$$

Ak sa objekt objavil na scéne až v kroku *t*, jeho percept $f^{(t-1)}$ bude mať špeciálnu hodnotu $\perp$. Ak bol objekt na scéne v kroku t - 1 a teraz zmizol, $f^{(t)} = \perp$.[7] Kritériá rozlišujúce zmenu vlastností objektu v čase sú realizované ako detektor aplikovaný na rámec $\Delta\left(f_1^{(t)}, f_1^{(t-1)}\right)$.

4.6 Sémantika slovies

Zachytenie dynamiky prostredia kritériami zmeny je dôležitou súčasťou sémantiky slovies. Niektoré zmeny sa dajú zachytiť čisto kvalitatívnymi reláciami, napr. *narástol* môže byť reprezentované ako sgn($\Delta\left(f_1^{(t)}, f_1^{(t-1)}\right)$.*veľkosť*) = +1, iné môžu obsahovať predpoklady o typických veľkostiach zmien, napr. kritériá pohybu *plazíSa*, *kráča*, *beží* sa môžu líšiť v priemerných hodnotách $\Delta\left(f_1^{(t)}, f_1^{(t-1)}\right)$.*pozícia*. Kritériá s nulovým znamienkom zmeny niektorých atribútov môžu reprezentovať stav, resp. trvanie vlastnosti, napr. *stojí*.

Komplexnejšia reprezentácia sémantiky slovies je prepojená s akciami a zahŕňa spôsoby ich vykonania, napr. reprezentácia bežania môže obsahovať prepojenie na nedeklaratívnu procedurálnu reprezentáciu invariantov motorického stereotypu spolu s rámcom variabilných parametrov akcie, napr. rýchlosti alebo natočenia končatín [1].

Keďže stelesnená reprezentácia významov vzniká v senzomotorických interakciách s prostredím, obsahuje aj situovanú kauzálnu znalosť o podmienkach úspechu akcie a jej možných dôsledkoch. Napríklad akcia *zdvihni* vykonaná rovnakým spôsobom (s rovnakou silou) môže viesť k rôznym výsledkom podľa toho, či zdvíhaným objektom je pingpongová loptička alebo 200 kilový kus nábytku. Agent preto využíva tie isté

[7] Na detekovaní $\perp$ na príslušnej pozícii vstupného páru sú založené špeciálne kritériá zmeny *objavilSa* a *zmizol*.

reprezentačné štruktúry na porozumenie významu slov, usudzovanie a plánovanie vlastných akcií v prostredí.

Vhodnou reprezentačnou štruktúrou pre zachytenie propozičných znalostí sú asociácie typu (*podmienky, akcia, dôsledky*), kde *podmienky* sú okolnosti (alebo objekt) akcie reprezentované kritériami situácií (alebo objektu), *akcia* kritériom reprezentujúcim typ a parametre vykonávanej akcie a *dôsledky* kritériami zmien, ktoré má akcia spôsobiť [33]. BDI[8] agent [6] môže reprezentovať svoje ciele ako kritériá želaných situácií a plánovať sekvencie akcií, ktoré by mohli viesť z aktuálnej situácie do cieľovej.

Takáto reprezentácia je v súlade s poznatkami o ľudskej organizácii reprezentácie znalostí. Podľa Gibsona [13] dokážeme v prostredí priamo vnímať možnosti akcií a manipulovania s objektmi: po zemi sa dá chodiť, dvere sa dajú otvoriť, jablko sa dá zjesť, atď. Vnímateľné interakčné vlastnosti predmetov Gibson nazval *afordanciami*. V našom modeli im zodpovedajú dôsledky asociované s objektovými kategóriami. Podľa Tomasella [36] je mentálny lexikón dieťaťa spočiatku organizovaný okolo konkrétnych slovies s konkrétnymi scenármi vykonávania. Takéto štruktúry Tomasello nazval *slovesnými ostrovmi*. V našom modeli slovesné ostrovy tvoria objekty a dôsledky asociované s akciami.

5 Osvojovanie si kritérií

V predošlom texte sme opísali, ako môže agent odvodiť rozlišovacie kritérium nejakého konceptu indukciou z množiny jeho príkladov. V skutočnosti však agent vníma zmes rámcov, ktoré sú príkladmi rôznych kategórií a musí určiť, ktoré z existujúcich kritérií má pomocou perceptu aktualizovať (alebo či má vytvoriť nové kritérium).

Združovanie do kategórií môže prebiehať rôznymi spôsobmi, podľa dostupnosti spätnej väzby z prostredia. Ak agent dostáva iba vstupy z prostredia, ale nemá žiadnu spätnoväzobnú slučku, môže prirodzeným spôsobom vytvoriť kategórie tak, aby odrážali štatistické rozdelenie vlastností prostredia. Prirodzená kategorizácia maximalizuje medzikategoriálne rozdiely a minimalizuje rozdiely v rámci kategórie. Na vytvorenie prirodzenej kategorizácie možno použiť klasterizačné algoritmy ako *Distributed Clustering Algorithm* [19], *1-nearest neighbour* a iné [8].

Prirodzená kategorizácia nemusí byť pre ciele systému postačujúca: napr. živočích, ktorý sa živí hubami (a chce prežiť), sa musí naučiť vnímať jemné odlišnosti vo vzhľade aj veľmi podobných húb, pokiaľ niektoré sú jedlé a iné jedovaté. Tu je vhodným typom kategorizácia na základe pragmatickej spätnej väzby. V článku [33] opisujeme model, v ktorom agent aktívne interagoval s objektmi v prostredí a zoskupoval ich do kategórií na základe ich afordancií. Agent sa pokúšal objekty zdvíhať alebo položiť a objekty tvorili jednu kategóriu, ak vykonanie niektorej akcie s nimi viedlo k rovnakému výsledku, napr. kategóriu objektov „príliš ťažkých na to, aby sa dali zdvihnúť“.

Tretím typom je kategorizácia s pomocou jazykových inštrukcií externého učiteľa. Externým učiteľom agenta môžu byť ostatné agenty, ktoré už zdieľajú spoločný jazyk, alebo človek. Napríklad dieťa si môže vyformovať kategóriu objektov, ktoré jeho rodič pomenúva tým istým menom – takto sa deti učia väčšinu nadradených kategórií ako zviera, či nábytok [39]. Kategórie osvojené podľa inštrukcií v prirodzenom jazyku sú

[8] belief, desire, intention

kultúrne podmienené – známym príkladom je existencia rôznych kategorizácií farieb vo svetových jazykoch [4]. V ďalšej časti sa zameriame na formovanie významu pojmov pomocou jazyka. Model nemá zabudované žiadne kultúrne špecifické predpoklady a môže si rovnako dobre osvojovať kategórie z inštrukcií v umelom jazyku.[9] Proces formovania kritérií je takto spriahnutý s procesom osvojovania jazyka a my budeme skúmať ich vzájomnú interakciu.

6 Spriahnuté osvojovanie si jazyka a kategórií

Agent dostane v každom časovom kroku na vstupe scénu, spolu s ňou jazykový výraz opisujúci niektoré jej aspekty a tiež tzv. fókus. Fókus je neverbálna referencia upriamujúca pozornosť agenta na aspekty scény pomenované jazykovým výrazom. Pri osvojovaní si jazyka u detí zodpovedá ukazovaniu, sledovaniu pohľadu, či zdieľanej pozornosti dieťaťa s matkou [37]; v neskoršej fáze, keď už agent rozumie niektorým slovám, kontext referencie môže byť zúžený na základe známych častí výrazu, napr. výraz „veľký X“ môže zúžiť kontext pre X na veľké objekty na scéne. Formálne je fókus projekciou scény na k-ticu rámcov ($k = 1$ pre objektovú kategóriu, $k = 2$ pre vzťahovú kategóriu alebo kategóriu zmeny; v prípade kategórie zmeny je jeden z rámcov podobou objektu v predošlom časovom kroku). Výsledok projektívnej funkcie budeme nazývať *referentom* fókusu.

V najjednoduchšej učiacej situácii agent vníma scénu S_t, jednoslovný jazykový výraz w a fókus ϕ. Agentove jazykové znalosti sú ukladané vo forme tabuľky asociácií slov (postupností písmen) a ich významov. Významom slova je rozlišovacie kritérium odvodené indukciou zo všetkých kontextov (referentov), s ktorými bolo slovo použité.

Hoci prirodzené jazyky obsahujú slová s viacerými význammi (homonymá) a viaceré slová s rovnakým významom (synonymá), v prípade spriahnutej akvizície jazyka a konceptov bude veľmi užitočné začať bez synonymie a homonymie.[10] V skutočnosti práve toto sú hlavné princípy, na ktorých je v našom modeli založené osvojovanie:

1. *Žiadna pravá homonymia*. Každé slovo má iba jeden význam, aj keď je použité s viacerými referentmi (vtedy sa referenty pokladajú za inštancie/príklady tej istej kategórie).[11] V praxi sa tento princíp využíva tak, že ak agent už pozná nejaký význam slova w (má s ním asociované kritérium r) a teraz je slovo w použité s novým referentom $\phi(S_t)$, r je aktualizované pomocou $\phi(S_t)$ (namiesto vytvorenia nového kritéria).
2. *Žiadna pravá synonymia.* Rôzne slová majú rôzny význam, aj keď zdieľajú ten istý referent (v tom prípade označujú rôzne aspekty referentu). V praxi sa tento princíp využíva tak, že ak agent počuje neznáme slovo w v kontexte referentu $\phi(S_t)$, vytvorí sa nové kritérium na základe prvého príkladu $\phi(S_t)$ a v lexikóne sa asociuje so slovom w.

[9] Jediným zabudovaným predpokladom je konvexnosť prirodzených kategórií.

[10] Deti osvojujúce si jazyk používajú *predpoklad vzájomnej výlučnosti* pomenovaní, teda predpoklad, že nové slová nemôžu označovať objekty, ktoré už majú nejaké iné meno [25]. Inou formuláciou je *princíp kontrastu* [7], ktorý predpokladá, že každej odlišnosti vo forme zodpovedá odlišnosť vo význame.

[11] Ak bolo slovo použité v zjavne úplne odlišných kontextoch (napr. referenty nemajú nič spoločné), agent by mohol detekovať homonymiu a asociovať so slovom viaceré kritériá; toto však v našom modeli nie je implementované.

Príklad: Uvažujme agenta, ktorý žije vo svete geometrických tvarov na mriežke 50×50 so súradnicami (1, 1) v ľavom dolnom a (50, 50) v pravom hornom rohu. Ak agent vníma percept

```
f = { vertices: 3; size: 18; color: 3; posX: 1; posY: 23 }
```

označený slovami „vľavo“, „veľký“, „trojuholník“, vytvorí si tri rozlišovacie kritériá, ktoré sú na začiatku identické a reprezentujú „fotku“ percipovaného objektu. Postupom času agent počuje tieto slová s ďalšími objektmi a kritériá sa začínajú diferencovať. Napr. počiatočná „fotka“ asociovaná so slovom „trojuholník“ sa aktualizuje pomocou perceptuálnych rámcov rôznych objektov najrôznejších farieb, pozícií, veľkostí a iných vlastností. Rámce všetkých objektov však budú mať spoločné to, že majú tri vrcholy. Atribúty, ktoré sa nevyskytujú v každom príklade, sa z kritéria vylúčia, iné s veľkým rozptylom budú mať malú váhu. Rozhodujúcou (a nutnou) pre príslušnosť v kategórii asociovanej so slovom „trojuholník“ sa stane vlastnosť mať tri vrcholy (keďže počet vrcholov má vo vzorke priemernú hodnotu 3 a rozptyl nula). Takisto slovo „vľavo“ bude agent počuť v kontexte najrôznejších objektov, ktoré všetky budú mať veľmi malú hodnotu atribútu *posX*, atď. Čím viac je rôznych kontextov použitia slova, tým väčšia je pravdepodobnosť, že referenty sa budú líšiť v hodnotách atribútov nepodstatných pre význam slova. Avšak keby boli napr. všetky trojuholníky v agentovom svete veľké, vlastnosť byť veľkým by sa stala súčasťou významu slova „trojuholník“.

Princípy používané pri osvojovaní si významov jednotlivých slov sa dajú aplikovať aj na viacslovné menné frázy: predpokladá sa, že každé slovo vo fráze označuje iný aspekt referentu. V modeloch akvizície bez gramatiky je poradie slov vo fráze nepodstatné a indukcia významov z frázy „veľký trojuholník vľavo“ (alebo jej ľubovoľnej permutácie) má rovnaký efekt ako tri postupné indukcie z jednotlivých slov tak, ako to bolo opísané v príklade v predchádzajúcej časti.

V reálnych gramatických jazykoch môže rôzne poradie slov dávať celej fráze rôzne významy.[12] Takisto význam adjektív v mennej fráze môže byť kontextovo závislý od významu podstatného mena, napr. slovo „malý“ označuje rôzne veľkosti vo frázach „malý zajac“ a „malý slon“. Proces osvojovania, tak ako sme ho opísali vyššie, by jednoducho aktualizoval kritérium asociované so slovom „malý“ pomocou perceptov slona a zajaca, čím by sme dostali kritérium s priemernou veľkosťou medzi zajacom a slonom a veľmi veľkým rozptylom. Rozumnejšie by bolo vytvoriť pre každý menný kontext iné kritérium pre „malý“ a aktualizovať každé zvlášť, to by si však už od agenta vyžadovalo istú znalosť gramatickej štruktúry osvojovaného jazyka (na rozlíšenie, ktorá časť mennej frázy je podstatné meno, a ktoré sú rozvíjajúce adjektíva).

Bez znalosti pravidiel gramatickej kompozície agent nemôže pochopiť ani význam fráz ako „mačka na horúcej plechovej streche“, aj keby si jednotlivo osvojil správne významy slov „mačka“, „na“, „horúca“, „plechová“, „strecha“.[13]

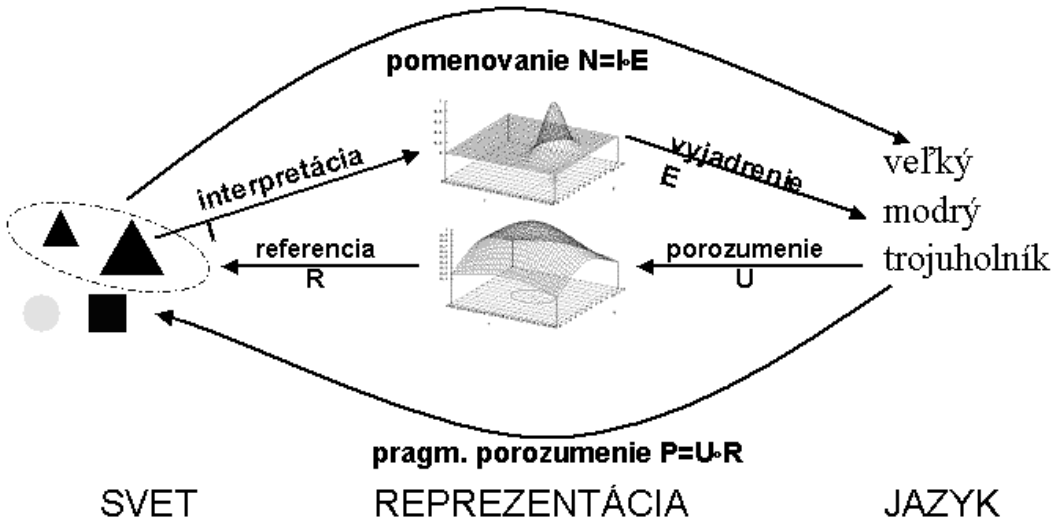
---

[12] Deti pri osvojovaní jazyka využívajú nielen poradie slov, ale aj ďalšie gramatické informácie na obmedzenie možných významov neznámych častí frázy (teória *syntaktického bootstrappingu* [24]). Obohatenie nášho modelu o citlivosť na gramatickú štruktúru je dôležitou témou pre budúci výskum.

[13] V slovenčine sú ďalšou komplikáciou, od ktorej abstrahujeme, rôzne gramatické tvary toho istého slova, napr. „horúca“ a„horúcej“. Avšak ani v angličtine, kde sú tvary rovnaké, na určenie významu frázy „cat on a hot tin roof“ nestačí poznať významy jednotlivých slov (bez znalosti ich vzťahov daných gramatikou by sme význam uvedenej frázy nevedeli odlíšiť napr. od významu fráz „hot tin cat on a roof“ alebo „tin roof on a hot cat“).

7 Pragmatická úroveň

Opísali sme mechanizmy, akými si agent môže interakciou s prostredím vybudovať rozlišovacie kritériá reprezentujúce rôzne koncepty a ukotviť v nich významy jazykových výrazov. To, ako bude agent naučené kritériá a osvojené jazykové výrazy ďalej používať, závisí od konkrétnej aplikácie. Pre komunikáciu je však užitočné definovať niekoľko všeobecných pragmatických funkcií zabezpečujúcich zobrazovanie medzi prostredím, repertoárom významov agenta a jazykom (obr. 4). Význam jednotlivých funkcií by mal byť z obrázku zrejmý, ich formálne definície čitateľ nájde v [34]. Dodáme len, že manipuláciou s rozhodovacím prahom θ určujúcim receptívne polia kritérií dosiahneme porozumenie kontrastívnemu použitiu slov. Ak napríklad agent počuje slovo „veľký“ so scénou obsahujúcou iba malé objekty, znížením prahu kritérií mu funkcia referencie vráti objekty, na ktoré sa popis najlepšie hodí, teda najväčšie z malých objektov prítomných na scéne.



Obrázok 4. Pragmatické funkcie zabezpečujú prepojenie prostredia, vnútorných reprezentácií a jazyka.

8 Záver

Rozlišovacie kritériá poskytujú jednotný reprezentačný formalizmus pre koncepty rôznych typov a ich osvojovanie. Proces formovania kritérií je ukotvený v percepcii a tesne spriahnutý s jazykom, čím sa navrhnutý formalizmus stáva dobrým kandidátom na kognitívnu sémantiku jazyka. Keďže kritériá sú citlivé na rozptyl a korelácie hodnôt atribútov v rámci kategórie, sú flexibilnejšou a efektívnejšou reprezentáciou ako „čisté“[14] prototypy v konceptuálnom priestore, čo sme aj experimentálne ukázali v našom implementovanom modeli interakčného osvojovania kritérií [33]. Rozlišovacie kritériá sú implementované pomocou lokálne reagujúcich detektorov a ich hierarchických sietí, čo je biologicky plauzibilné.

Ak by sme chceli pomocou formalizmu rozlišovacích kritérií realistickejšie reprezentovať sémantiku slovies, bolo by ho treba rozšíriť o štruktúry reprezentujúce

[14] S euklidovskou vzdialenosťou.

perspektívu, spôsob, roly, aspekt a ďalšie parametre slovies. Indukčný proces v súčasnej podobe nie je robustný voči chybám a šumu (jeden nesprávny príklad môže nevratne ovplyvniť množinu atribútov kritéria), preto by ho bolo rozumné nahradiť stochastickou indukciou. V budúcnosti tiež treba dopracovať detaily kritérií situácií a ich indukciu a rozšíriť jazykovú úroveň o viacslovné frázy a nejakú formu gramatiky.

V každom prípade považujeme navrhnutú sémantiku za solídnu základňu pre kognitívne modely osvojovania jazyka, BDI architektúry a aplikácie v otvorených umelých prostrediach, akým je napr. world wide web.

Poďakovanie: Tento výskum bol financovaný z grantu č. APVV-20P04805 Agentúry na podporu výskumu a vývoja a z grantu VEGA 1/3105/06.

Literatúra

[1] Bailey, D., Feldman, J., Narayanan, S., Lakoff, G.: Modeling Embodied Lexical Development. In *Proceedings of the 19th Cognitive Science Society Conference* (pp. 19–24) 1997.

[2] Balkenius, C.: Are There Dimensions in the Brain? In *Spinning Ideas, Electronic Essays Dedicated to Peter Gärdenfors on his Fiftieth Birthday*. 1999. http://www.lucs.lu.se/spinning/categories/cognitive/Balkenius/Balkenius.pdf

[3] Barsalou, L. W.: Perceptual Symbols Systems. *Behavioral and Brain Sciences* **22** (1999) 577–660.

[4] Berlin, B., Kay, P.: *Basic Color Terms: Their Universality and Evolution*. Berkeley, CA: University of California Press 1969.

[5] Bloom, P.: *How Children Learn the Meanings of Words*. Cambridge, MA: MIT Press 2000.

[6] Bratman, M.: *Intention, Plans, and Practical Reason*. Cambridge, MA: Harvard University Press 1987.

[7] Clark, E. V.: The Principle of Contrast: A Constraint on Language Acquisition. In MacWhinney, B. (Ed.): *Mechanisms of Language Acquisition*. London: Erlbaum 1987.

[8] Everitt, B. S., Landau, S., Leese, M.: *Cluster Analysis*. London: Arnold 2001.

[9] Feldman, J.: *From Molecule to Metaphor: A Neural Theory of Language*. Cambridge, MA: MIT Press 2006.

[10] Fodor, J.: *Representations: Philosophical Essays on the Foundations of Cognitive Science*. Cambridge, MA: MIT Press 1981.

[11] Gärdenfors, P.: *Conceptual Spaces*. Cambridge, MA: MIT Press 2000.

[12] Geuder, W., Weisgerber, M.: Verbs in Conceptual Space. In Katz, G., Reinhard, S., Reuter, P. (Eds.): *Sinn und Bedeutung 6, Proceedings of the Sixth Meeting of the Gesellschaft für Semantik*, volume 1, Osnabrück: Publications of the Institute of Cognitive Science 2002.

[13] Gibson, J. J.: *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin 1979.

[14] Gold, E. M.: Language Identification in the Limit. *Information and Control* **10**(5) (1967) 447–474.

[15] Gopnik, A., Meltzoff, A. N., Kuhl, P. K.: *The Scientist in the Crib: What Early Learning Tells Us About the Mind*. New York: HarperCollins 2000.
[16] Harnad, S.: The Symbol Grounding Problem. *Physica D* **42** (1990) 335–346.
[17] Hassoun, M. H.: *Fundamentals of Artificial Neural Networks*. Cambridge, MA: MIT Press 1995.
[18] Haykin, S.: *Neural Networks: A Comprehensive Foundation (Second ed.)*. Upper Saddle River, NJ: Prentice Hall 1999.
[19] Hulth, N., Grenholm, P.: A Distributed Clustering Algorithm. *Lund University Cognitive Studies* **74** (1998).
[20] Ientilucci, E. J.: Using the Singular Value Decomposition. 2003. http://www.cis.rit.edu/~ejipci/Reports/svd.pdf
[21] Johnson, S. P., Amso, D., Slemmer, J. A.: Development of Object Concepts in Infancy: Evidence for Early Learning in an Eye Tracking Paradigm. *Proceedings of the National Academy of Sciences USA* **100** (2003) 10568–10573.
[22] Kirby, S., Hurford, J.: The Emergence of Linguistic Structure: an Overview of the Iterated Learning Model. In Parisi, D., Cangelosi, A. (Eds.): *Computational Approaches to the Evolution of Language and Communication*. Berlin: Springer-Verlag 2001.
[23] Langacker, R.: *Concept, Image and Symbol: The Cognitive Basis of Grammar*. Mouton de Gruyter 1991.
[24] Lidz, J., Gleitman, H., Gleitman, L. R.: Kidz in the ’Hood: Syntactic Bootstrapping and the Mental Lexicon. In Hall, D. G., Waxman, S. R. (Eds.): *Weaving a Lexicon.* Cambridge, MA: MIT Press 2004.
[25] Markman, E.: Constraints on Word Learning: Speculations About Their Origins and Domain Specificity. In Gunnar, M. R., Maratsos, M. (Eds.): *Modularity and Constraints in Language and Cognition*. Hillsdale, NJ: Lawrence Erlbaum Associates 1992.
[26] Martin, J. H.: Coding and Processing of Sensory Information. In Kandel, E. R., Schwartz, J. H., Jessel, T. M. (Eds.): *Principles of Neural Science* (pp. 329–340). New York: Elsevier 1991.
[27] Medin, D. L., Altom, M. W., Edelson, S. M., Freko, D.: Correlated Symptoms and Simulated Medical Classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **8** (1982) 37–50.
[28] Piaget, J., Inhelder, B.: *La Psychologie de L’enfant*. Paris: PUF 1966.
[29] Spelke, E. S.: Principles of Object Perception. *Cognitive Science* **14** (1990) 29–56.
[30] Steels, L.: Language as a Complex Adaptive System. In Schoenauer, M. (Ed.): *Proceedings of PPSN-VI*, Berlin: Springer-Verlag 2000.
[31] Šefránek, J.: Kognícia bez mentálnych procesov. In Beňušková, Ľ. et al. (Eds.): *Kognitívne vedy.* Bratislava: Kaligram 2002.
[32] Takáč, M.: *Kvalitatívne modelovanie a simulácia*. Bratislava: Vydavateľstvo UK 2003.
[33] Takáč, M.: 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.* Trieste: Edizione Goliardiche 2006.
[34] Takáč, M.: Kognitívna sémantika rozlišovacích kritérií. In J. Kelemen, V. Kvasnička (Eds.): *Kognice a umělý život VI*. Opava: Slezská univerzita 2006.

[35] Talmy, L.: *Toward a Cognitive Semantics*. Cambridge, MA: MIT Press 2000.
[36] Tomasello, M.: *First Verbs: A Case Study of Early Grammatical Development*. Cambridge: CUP 1992.
[37] Tomasello, M., Farrar, J.: Joint Attention and Early Language. *Child Development* **57** (1986) 1454– 1463.
[38] Tversky, A.: Features of Similarity. *Psychological Review* **84** (4) (1977) 327–352.
[39] Waxman, S. R., Braun, I.: Consistent (but not Variable) Names as Invitations to Form Object Categories: New Evidence from 12-Month-Old Infants. *Cognition* **95** (2005) B59-B68.
[40] Whorf, B. L.: *Language, Thought and Reality: Selected writings of Benjamin Lee Whorf*. Cambridge, MA: MIT Press 1956.

Korektúra textu neurónovými sieťami s echo stavmi

Peter Trebatický[1]

Abstrakt. Veľa úsilia pri vytváraní rozpoznávacích systémov je venované korekcii poškodených postupností. Vo veľa systémoch na rozpoznanie optických a ručne písaných znakov je posledným krokom korektúra textu použitím techník analýzy prirodzeného jazyka. Neurónové siete s echo stavmi predstavujú nový pohľad na trénovanie rekurentných neurónových sietí. Skúmame možnosť použiť tieto siete, ktoré sa jednoducho trénujú, na korekciu textu z anglického prekladu biblie (King James) a porovnávame s výkonom Markovovského modelu s premenlivou dĺžkou.

1 Úvod

V tomto článku sa zaoberáme možnosťou využitia rekurentných neurónových sietí s echo stavmi na korekciu poškodeného textu. Keďže tieto siete majú vysoký stupeň biologickej plauzability a ich trénovanie je pomerne jednoduché, má význam sa touto problematikou bližšie zaoberať. Veľa úsilia pri vytváraní rozpoznávacích systémov je venované korekcii poškodených postupností. Vo veľa systémoch na rozpoznanie optických a ručne písaných znakov je posledným krokom korektúra textu použitím techník analýzy prirodzeného jazyka.

V nasledujúcej kapitole čitateľa stručne uvedieme do problematiky neurónových sietí s echo stavmi (Echo State Networks – ESN) a popíšeme základnú techniku trénovania týchto pomerne nových sietí. Pre lepšie pochopenie ich princípu sme ako tretiu sekciu zaradili tzv. automaty s kvapalinovými stavmi, ktoré sú všeobecnejšie a zahŕňajú aj neurónové siete s echo stavmi. Potom nasleduje popis Markovovských modelov s premenlivou dĺžkou, ktoré už boli na úlohu korekcie použité [7]. Po vysvetlení techniky použitej na túto úlohu nasleduje vysvetlenie ako ju aplikovať na Markovovské modely a tiež na neurónové siete s echo stavmi. Nakoniec nasleduje popis experimentov, porovnanie výsledkov oboch techník.

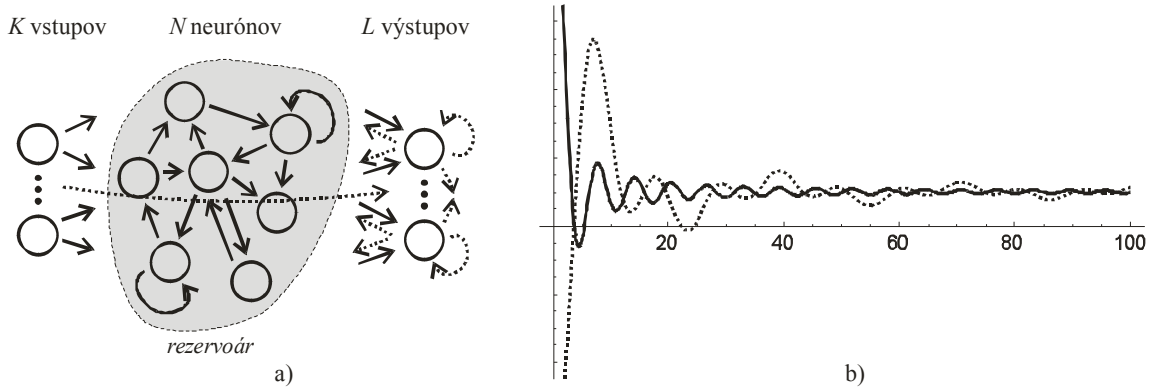
2 Neurónové siete s echo stavmi

Rekurentné neurónové siete s echo stavmi predstavujú nový pohľad na trénovanie neurónových sietí. Základnou myšlienkou je použitie veľkého rezervoára rekurentne poprepájaných neurónov ako zdroj zaujímavých signálov, z ktorých sa vyskladá požadovaný výstup. Čo je pre neurónové siete s echo stavmi oproti klasickým technikám „netradičné", váhy spojení v tomto rezervoári sa počas trénovania nemenia, menia sa len výstupné váhy (čo zodpovedá „skladaniu" požadovaného výstupu) [2]. Tento prístup úzko súvisí s tzv. architekturálnym biasom [9], kedy aj nenatrénovaná sieť má užitočné vlastnosti.

Predpokladáme neurónovú sieť pracujúcu v diskrétnom čase s K vstupnými neurónmi, N internými neurónmi a L výstupnými neurónmi. Aktivácie vstupných neurónov v čase n sú $\mathbf{u}(n) = (u_1(n),\ldots,u_K(n))$, skrytých neurónov $\mathbf{x}(n) = (x_1(n),\ldots,x_N(n))$

[1] Fakulta informatiky a informačných technológií, STU, 842 13 Bratislava, E-mail: trebaticky@fiit.stuba.sk

a výstupných neurónov $\mathbf{y}(n) = (y_1(n), \ldots, y_L(n))$. Matica $\mathbf{W}^{in}_{[N\times K]} = (w^{in}_{ij})$ pozostáva z váh vstupných spojení, matica $\mathbf{W}_{[N\times N]} = (w_{ij})$ z váh vnútorných spojení, matica $\mathbf{W}^{out}_{[N\times(K+N+L)]} = (w^{out}_{ij})$ z váh spojení k výstupom a matica $\mathbf{W}^{back}_{[N\times L]} = (w^{back}_{ij})$ z váh spätnoväzobných spojení z výstupov ku skrytým neurónom [3]. Na topológiu v rezervoári nie sú kladené žiadne požiadavky, t. j. nemusí mať vrstvovú štruktúru. Predpokladáme, že v rezervoári sú vytvorené rekurentné spojenia medzi skrytými neurónmi. Štruktúra neurónovej siete je na obrázku 1a.



Obrázok 1. a) Štruktúra neurónovej siete s echo stavmi. Čiarkované spojenia sú možné, ale nie sú vyžadované. b) Ilustrácia vlastnosti siete s echo stavmi – stav siete po dostatočne dlhom čase pre rovnakú vstupnú postupnosť nezávisí od počiatočného stavu.

Výstupy skrytých neurónov sú aktualizované na základe rovnice

$$\mathbf{x}(n+1) = \mathrm{f}(\mathbf{W}^{in}\mathbf{u}(n+1) + \mathbf{W}\mathbf{x}(n) + \mathbf{W}^{back}\mathbf{y}(n)) \tag{1}$$

kde $\mathrm{f} = (f_1, \ldots, f_N)$ sú výstupné funkcie skrytých neurónov (typicky sigmoida). Výstup sa vypočíta nasledovne

$$\mathbf{y}(n+1) = \mathrm{f}^{out}(\mathbf{W}^{out}(\mathbf{u}(n+1) \quad \mathbf{x}(n+1) \quad \mathbf{y}(n))) \tag{2}$$

kde $\mathrm{f}^{out} = (f^{out}_1, \ldots, f^{out}_L)$ sú výstupné funkcie výstupných neurónov a $(\boldsymbol{u}(n+1) \quad \boldsymbol{x}(n+1) \quad \boldsymbol{y}(n))$ predstavuje spojenie vstupného, vnútorného a výstupného vektora do jedného.

Neurónová sieť má echo stavy, keď sa pre rovnakú vstupnú sekvenciu ustáli (po nekonečnom čase) na rovnakom stave bez ohľadu na počiatočný stav siete [1] (obr. 1b).

Na to, aby neurónová sieť mala echo stavy je potrebné, aby matica váh $\mathbf{W}$ mala určité vlastnosti [3]. Začneme s nejakou maticou $\widetilde{\mathbf{W}}$ a prenásobíme ju nejakým vhodným škálovacím faktorom α. Nech $\sigma_{\max}(\mathbf{W})$ predstavuje najväčšiu singulárnu hodnotu matice $\mathbf{W}$ a $|\lambda_{\max}|(\mathbf{W})$ spektrálny rádius matice $\mathbf{W}$. Ak $\alpha_{\min} = \sigma_{\max}(\widetilde{\mathbf{W}})$ a $\alpha_{\max} = 1/|\lambda_{\max}|(\widetilde{\mathbf{W}})$, potom ak zvolíme faktor $\alpha < \alpha_{\min}$, sieť bude určite mať echo stavy, a ak $\alpha > \alpha_{\max}$, potom sieť echo stavy určite nemá. V praxi sa však ukazuje, že sieť má echo stavy aj ak je α len marginálne menšia ako $\alpha_{\max}$.

Základný rozdiel oproti „klasickým" trénovacím technikám neurónových sietí spočíva v tom, že váhy spojení medzi skrytými neurónmi sa v trénovacom procese nemenia, t. j. matica $\mathbf{W}$ sa po škálovacom kroku už neaktualizuje. Podstata trénovania totiž spočíva v skombinovaní signálov od jednotlivých skrytých neurónov tak, aby sa skutočný výstup čo najviac priblížil k požadovanému, t. j. minimalizovať chybu štvorcov. Na to môžeme použiť ľubovoľnú metódu lineárnej regresie. Trénovaním teda zmeníme len váhy spojení k výstupným neurónom – maticu $\mathbf{W}^{out}$.

Trénovanie siete s echo stavmi (t. j. s vhodnou maticou $\mathbf{W}$) vo všeobecnosti prebieha v nasledujúcich krokoch [3]:

1. Zostrojíme ľubovoľné matice $\mathbf{W}^{in}_{[N\times K]}$ a $\mathbf{W}^{back}_{[N\times L]}$.
2. Začneme z ľubovoľného stavu siete $\mathbf{x}(0)$ a aktualizujeme ho v časoch $n = 0,\ldots,n_{\max}$:
$$\mathbf{x}(n+1) = \mathrm{f}(\mathbf{W}^{in}\mathbf{u}(n+1) + \mathbf{W}\mathbf{x}(n) + \mathbf{W}^{back}\mathbf{y}(n)) \tag{3}$$
Odstránime dostatočný počet krokov $n_{\min}$, aby sme minimalizovali vplyv počiatočného stavu. Z týchto stavov zostrojíme maticu $\mathbf{X}_{[N\times M]}$, kde $M = n_{\max} - n_{\min}$.
3. Vypočítame maticu váh $\mathbf{W}^{out}_{[L\times N]}$ tak, aby sme minimalizovali trénovaciu chybu. Nech $\mathbf{y}_{teach}(n) = (y_{teach,1}(n),\ldots,y_{teach,L}(n))$ a nech $g_j(n) = (f_j^{out})^{-1} y_{teach,j}(n)$. Pre všetky $j = 1,\ldots,L$ vypočítame váhy w_{ji}^{out} tak, aby sme minimalizovali strednú kvadratickú chybu:
$$\frac{1}{n_{\max} - n_{\min}} \sum_{n=n_{\min}}^{n_{\max}} \left(g_j(n) - \sum_{i=1}^{N} w_{ji}^{out} x_i(n) \right)^2 \tag{4}$$
Na tento cieľ môžeme použiť ľubovoľnú metódu lineárnej regresie. Preto použijeme napr. nasledovnú metódu. Z upravených požadovaných výstupov $\mathbf{g}(n_{\min}+1),\ldots,\mathbf{g}(n_{\max})$ zostrojíme maticu $\mathbf{Y}_{[L\times M]}$. Máme úlohu vyriešiť maticovú rovnicu
$$\mathbf{W}^{out}\mathbf{X} = \mathbf{Y} \tag{5}$$
pre neznámu $\mathbf{W}^{out}$. Jej riešením je
$$\mathbf{W}^{out} = \mathbf{Y}\left[\left(\mathbf{X}^T\right)^{-1}\right]^T \tag{6}$$
kde výraz $\mathbf{A}^{-1}$ predstavuje pseudoinverznú maticu matice $\mathbf{A}$. Po vypočítaní matice $\mathbf{W}^{out}$ je sieť natrénovaná.

Pri dlhých vstupných sekvenciách je však výpočet pseudoinverznej matice výpočtovo náročný, preto je v takomto prípade vhodnejšie použiť iteračnú metódu na lineárnu regresiu – napríklad Recursive Least Squares (RLS) [8]:
$$\begin{aligned}
\mathbf{k}(n) &= \frac{\mathbf{P}(n-1)\mathbf{x}(n)}{\lambda + \mathbf{x}^T(n)\mathbf{P}(n-1)\mathbf{x}(n)} \\
\boldsymbol{\alpha}(n) &= \mathbf{d}(n) - \mathbf{x}^T(n)\mathbf{W}^{out}(n-1) \\
\mathbf{W}^{out}(n) &= \mathbf{W}^{out}(n-1) + \mathbf{k}(n)\boldsymbol{\alpha}^T(n) \\
\mathbf{P}(n) &= \tfrac{1}{\lambda}(\mathbf{P}(n-1) - \mathbf{k}(n)\mathbf{x}^T(n)\mathbf{P}(n-1))
\end{aligned} \tag{7}$$

kde $\mathbf{d}(n)$ je požadovaný výstup v čase n, $\mathbf{P}(0) = b\ \mathbf{I}$, kde b>100, ale na tejto hodnote pri dlhých postupnostiach nezáleží a λ je tzv. zabúdací faktor medzi 0 a 1.

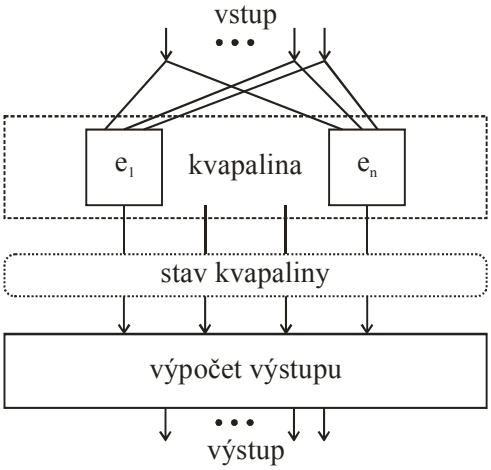
Ak nepoužívame vstupný signál, resp. spätnú väzbu, potom z rovnice 3 odstránime súčin obsahujúci odpovedajúcu premennú – $\mathbf{W}^{in}$, resp. $\mathbf{W}^{back}$.

Neurónové siete s echo stavmi boli síce vytvorené z matematického a inžinierskeho pohľadu, v skutočnosti však v sebe zahŕňajú typické znaky biologických neurónových sietí: veľký počet neurónov, rekurentné spojenia, riedke náhodné prepojenie a lokálne modifikovanie synaptických váh [3].

3 Automat s kvapalinovými stavmi

Neurónové siete s echo stavmi [6] sa v súčasnosti ukazujú byť špeciálnym prípadom tzv. Liquid State Machines (LSM) – automaty s kvapalinovými stavmi (obr. 3). Tieto sú založené na myšlienke, že teoreticky je možné z aktuálneho stavu povrchu hladiny (napr. v šálke) vyčítať, aké vplyvy na ňu pôsobili pred pozorovaním – vhodenie cukru, ponorenie lyžičky, apod. Inými slovami, aktuálny stav systému odráža históriu vstupov do systému. Výstup získaný pozorovaním tohto systému by mohla byť klasifikácia vplyvov nedávno pred pozorovaním, alebo predikcia ďalších očakávaných vplyvov.

Dôležité je si uvedomiť, že na zistenie predošlých stavov nie je potrebné uchovávať si predošlé stavy, a teda je možné počítať v reálnom čase. Samozrejme, aby toto fungovalo, aktuálny stav musí zahŕňať všetky relevantné informácie o vstupe a takisto dve rozdielne časové postupnosti vstupov by mali vyústiť do dvoch dostatočne rozdielnych stavov [6].



Obrázok 3. Automat s kvapalinovými stavmi. Vstupy ovplyvňujú elementy kvapaliny ei, ktoré tvoria jej stav. Výpočet výstupu je vytvorený tak, aby výstup zo systému čo najlepšie zodpovedal požadovanému [6].

Z tohto pohľadu môžeme počítanie výstupu chápať ako zariadenie, ktoré získava požadovanú informáciu o vstupe z aktuálneho stavu kvapaliny [6]. Táto úloha sa zjavne značne zjednoduší, ak stavy kvapaliny obsahujú v „jasne viditeľnej“ forme informáciu o vstupoch, ktoré máme pri získavaní výstupu zachytiť. Táto vlastnosť kvapaliny závisí od jej vnútornej štruktúry, najmä od typu a rozsahu interakcií medzi jednotlivými prvkami, z ktorých kvapalina pozostáva. Vnútorná štruktúra skutočných kvapalín je však nanešťastie taká, že má pomerne malé časové konštanty na odznenie účinkov vplyvov a

tieto vplyvy majú príliš lokálny vplyv. Preto napríklad šálka kávy nie je príliš vhodná ako výpočtový prostriedok.

Na druhej strane neurónové obvody by mohli byť vhodné „kvapaliny“, nakoľko v nich dochádza k distribuovaným interakciám medzi ich elementmi (t. j. neurónmi) a je v nich obsiahnutá veľká rôznorodosť dynamických sekvencií [6]. Tieto charakteristické vlastnosti potenciálne umožňujú neurálnym systémom fungovať ako optimálny integrátor všetkých druhov senzorických vstupov. Optimálny v tom zmysle, že stav neurálneho systému slúži ako univerzálny zdroj informácií o aktuálnom a minulých stavoch, ktoré je možné získať jednoduchým mechanizmom (napr. jediným neurónom) na vypočítanie požadovaného výstupu.

Podľa autorov [6] je jeden z empiricky najúspešnejších prístupov v strojovom učení sa založený na pozorovaní, že takmer všetky prakticky relevantné klasifikačné úlohy môžu byť vykonané jediným perceptrónom za predpokladu, že sú vstupy namapované na vysoko dimenzionálny priestor.

Autori uvádzajú nasledovný všeobecný postup pri implementácii LSM pre konkrétnu úlohu:

1. Zvoliť si vhodnú „kvapalinu“.
2. Uchovať si pozorovania stavu kvapaliny v rôznych časových okamihoch ako reakcie na viaceré rozdielne (trénovacie) vstupy.
3. Aplikovať algoritmus učenia s učiteľom na trénovaciu množinu tvorenú dvojicami (vstup, stav) získanú v kroku 2 tak, aby sa skutočné výstupy zo systému čo najviac blížili požadovaným výstupom.

Tieto všeobecné kroky zodpovedajú postupu pri trénovaní siete s echo stavmi popísanom v druhej kapitole. Prínosom autorov neurónových sietí s echo stavmi však je najmä to, že stanovujú konkrétne požiadavky na vlastnosti „kvapaliny“ (v ich prípade sigmoidálna rekurentná sieť) tak, aby mala požadované vlastnosti – t. j. aby bolo možné vyčítať históriu vstupov. Naproti tomu autori LSM v konečnom dôsledku len uvádzajú, aké sú to tie vlastnosti.

Nedávne zistenia v neurofyziológii naznačujú, že základný princíp sietí s echo stavmi, resp. automatov s kvapalinovými stavmi sa nezdá byť v biologických sieťach zriedkavý a mohol by byť eventuálne použitý pri riadení zariadení v protézach prostredníctvom signálov získaných zo zväzkov neurónov [4].

4 Markovovský model s premenlivou dĺžkou

Markovovské modely patria medzi stochastické metódy, kedy pravdepodobnosť nasledujúceho symbolu je určená na základe aktuálneho kontextu. Pri vytváraní Markovovského modelu s premenlivou dĺžkou [5, 7] prechádzame trénovaciu postupnosť a pre každý nájdený kontext nájdeme koľkokrát sa v postupnosti nachádza a pre každý jeho výskyt zistíme aký symbol za ním nasledoval. Takto dostaneme pre daný kontext tabuľku početnosti výskytov jednotlivých symbolov vstupnej abecedy bezprostredne nasledujúcimi za kontextom. Dôležité je, že jednotlivé kontexty nemusia mať rovnakú dĺžku.

Pri predikcii nasledujúceho symbolu potom postupujeme tak, že nájdeme najdlhší známy kontext pred symbolom, ktorý máme predikovať a za nasledujúci symbol

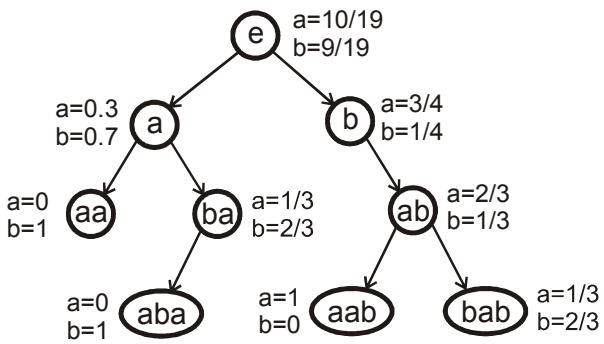
prehlásime symbol s najväčšou početnosťou pre daný kontext – vydelením počtom výskytov kontextu vo vstupnej postupnosti získame výsledný odhad pravdepodobnosti.

Aby sme vedeli rýchlo nájsť najdlhší kontext pre každú pozíciu vo vstupnom reťazci, vytvárame si tzv. Prediction Suffix Tree. Je to strom, ktorého koreňom je prázdny znak a jeho vrcholy zodpovedajú kontextom. V danej hĺbke stromu sú vrcholy kontextov rovnakej dĺžky. Dôležité je, že kontexty potomkov vrcholov zodpovedajú kontextu rodičovského vrcholu zľava rozšíreného o jeden symbol – preto názov suffix tree. Pri hľadaní kontextu potom postupujeme od koreňa stromu a pohybujeme sa v ňom hlbšie na základe posledného nespracovaného symbolu. Takto pokračujeme, kým sa nedostaneme do listu stromu, alebo keď už nie je žiadny nespracovaný symbol (čo sa môže stať len na začiatku postupnosti).

Pri vytváraní PST do stromu nezahŕňame kontexty, ktoré sa vo vstupnej postupnosti vyskytujú menej ako zvolený počet krát, prípadne by sme ich pridaním nezískali dodatočnú informáciu – ak rozdelenie pravdepodobnosti rodičovského vrcholu je blízke (menšie ako zvolený prah ε), čoho mierou je Kullback-Leibergova divergencia (1). Skúmajú sa však všetky kontexty do zvolenej hĺbky *L*. Príklad PST je na obr. 4.

$$Err(xw, w) = P(xw)\sum_{a \in A} P(a \mid xw) \log \frac{P(a \mid xw)}{P(a \mid w)} \tag{8}$$

kde x je symbol, xw je kandidát na nový kontext, w je kontext prislúchajúci rodičovskému uzlu v zostrojovanom PST a A je abeceda.



Obrázok 4. Príklad Markovovského modelu s premenlivou dĺžkou reprezentovaný štruktúrou Prediction Suffix Tree pre postupnosť aababbaabababbaabab

5 Viterbiho algoritmus

Oprava poškodeného textu (postupnosti symbolov) spočíva v nájdení najpravdepodobnejšej postupnosti stavov, ktorá vygenerovala pôvodný text [7]. Predpokladáme, že text bol vytvorený rovnakým stochastickým procesom (automatom), aký vytvoril trénovacie údaje, a že poznáme hodnotu šumu, ktorý na postupnosť vplýva.

Potom môžeme túto úlohu vyriešiť známym algoritmom dynamického programovania v odbore spracovania signálov, tzv. Viterbiho algoritmus [11]. Algoritmus môžeme popísať nasledovným programom v pseudo jazyku (tab. 1).

Tabuľka 1. Viterbiho algoritmus

SEQUENCE	poškodená postupnosť symbolov	*PS := počiatočné pravdepodobnosti stavov*
STATES	množina stavov PFA	
PS	vektor pravdepodobností, že sme v stave *i*	*foreach Sj from SEQUENCE do begin*
PSN	novovytváraný vektor *PS*	*PSN := nuly;*
ErrorMatrix	pravdepodobnosť zmeny symbolu *i* na *j*	*foreach T from STATES do begin*
		foreach N from Transitions(T) do begin
		K := NextState(N);
		P := ErrorMatrix [Sj, Symbol(N)];
Transitions(T)	množina prechodov zo stavu *T*	*if PSN [K] < P * PS [T] * Prob(N) then begin*
NextState(N)	nasledujúci stav zodpovedajúci prechodu *N*	*PSN [K] := P * PS [T] * Prob(N);*
		TabPrev [K,j] := T;
Symbol(N)	symbol zodpovedajúci prechodu *N*	*end;*
Prob(N)	pravdepodobnosť prechodu *N*	*end;*
		end;
TabPrev	predchádzajúci stav pre stav *i* a *j*-ty symbol v postupnosti	*PS := PSN;*
		end;
		ReconstructCorrectedSequence(TabPrev);

Na zrekonštruovanie opravenej postupnosti je potrebné nájsť maximálnu hodnotu vo vektore *PS*. Tak získame stav, v ktorom hľadaná najpravdepodobnejšia postupnosť končí. Predchádzajúci stav je v TabPrev[*K*, *j*], kde *K* je náš aktuálne získaný stav a *j* je pozícia symbolu v opravovanej postupnosti. Tento krok sa opakuje pre všetky symboly v postupnosti, pričom postupujeme odzadu.

Po získaní najpravdepodobnejšej postupnosti stavov vyberieme posledný symbol kontextu každého stavu, čím získame opravenú postupnosť symbolov. Celková časová zložitosť tohto algoritmu je $O(N\ T\ S)$, kde N je počet symbolov abecedy, T počet stavov a S dĺžka opravovanej postupnosti.

6 Transformovanie PST na PFA

Pred tým ako môžeme aplikovať techniku na korekciu textu je potrebné transformovať PST na konečný pravdepodobnostný automat (Probabilistic Finite Automaton – PFA). Na to je potrebné, aby všetky vnútorné vrcholy PST mali úplný stupeň a pre každý list s musí platiť, že jeho najdlhší prefix je buď list, alebo vnútorný vrchol výsledného PST. Pri pridávaní nových uzlov za účelom splnenia tejto požiadavky tieto vrcholy dedia rozdelenie pravdepodobnosti nasledujúceho symbolu od svojich rodičov. Stavy PFA sú tvorené listami výsledného PST vrátane rozdelenia pravdepodobnosti nasledujúceho symbolu [7].

Navyše je potrebné vypočítať rozdelenie pravdepodobnosti, že začneme v niektorom stave automatu. Nakoľko súčet týchto pravdepodobností je rovný jednej a počiatočná pravdepodobnosť každého stavu je rovná sume súčinov pravdepodobnosti prechodu do daného stavu a počiatočnej pravdepodobnosti stavu, z ktorého prechádzame, dostaneme n+1 rovníc o n neznámych. To vieme riešiť napríklad Gaussovou eliminačnou metódou.

7 Transformovanie ESN na PFA

Takisto ESN je potrebné transformovať na konečný pravdepodobnostný automat, aby sme mohli použiť techniku na opravu textu. V tejto časti popíšeme postup ako toho docieliť, ktorého myšlienka je založená na extrahovaní stochastických automatov z rekurentných neurónových sietí popísanom v [10].

Prvým krokom je kvantizovanie stavov rezervoára (t. j. aktivít rekurentných neurónov ako odozvy na vstupnú trénovaciu postupnosť) do klastrov. Použili sme techniku K-means, pričom sme kvôli výpočtovej náročnosti a pomalej konvergencii použili len časť zozbieraných stavov rezervoára – konkrétne každý 50. stav. Týmto získame centrá jednotlivých klastrov.

Každý klaster potom zodpovedá jednému stavu vytváraného automatu. Následne je potrebné nájsť pravdepodobnosti prechodov medzi týmito stavmi. Tieto odhadneme na základe relatívnych početností prechodov zo stavu prislúchajúcemu klastru i do stavu prislúchajúcemu klastru j. Dávame teda sieti na vstup trénovaciu postupnosť, sledujeme stavy rezervoára a zisťujeme príslušnosť každého stavu ku klastru (t. j. ku ktorému centru klastru je stav najbližšie v zmysle kvadrátu Euklidovskej vzdialenosti).

Nakoniec je potrebné pre každý stav automatu (klaster) zistiť rozdelenie pravdepodobnosti pozorovania jednotlivých symbolov v danom stave. To docielime opätovným prezentovaním trénovacej postupnosti a pozorovaním, na základe akého symbolu sme prešli do stavu rezervoára, ktorý zodpovedá danému stavu automatu (klastra). Takto pre každý stav automatu získame relatívne početnosti pozorovaných symbolov. Vo veľkej väčšine prípadov však výrazne vyčnieva práve jeden symbol, preto si k stavu automatu stačí pamätať iba tento jeden symbol, čím je možné použiť popísaný Viterbiho algoritmus.

8 Experimenty

Inšpirovaní článkom [7] sme ako trénovaciu množinu zvolili knihy King James biblie okrem knihy Genesis, ktorá slúžila ako testovacia množina. Trénovacia postupnosť obsahovala $2{,}9\times10^6$ symbolov a testovacia $1{,}9\times10^5$, pričom abeceda pozostávala z 27 rôznych znakov – malé písmená a medzera.

Na otestovanie predikčných schopností Markovovského modelu s premenlivou dĺžkou sme vytvorili Prediction Suffix Tree s parametrami $\varepsilon = 10^{-4}$ a $L = 30$, čím sme dostali 1790 vrcholov. Tento sme popísanou technikou transformovali na PFA s 13911 stavmi.

Pri tvorbe rekurentnej siete s echo stavmi sme použili rezervoár so 100 neurónmi. 20% zo všetkých možných rekurentných spojení bolo inicializovaných na rovnomerne náhodné desatinné číslo z intervalu (-1, 1), ostatné spojenia mali váhu nulovú. Maticu vnútorných spojení sme naškálovali tak, aby bol jej spektrálny rádius blízky 1, konkrétne 0,998. Je tomu tak preto, aby sme dosiahli dlhšiu krátkodobú pamäť. Vstupné váhy boli rovnomerne náhodne vybraté z celých čísel z intervalu <-5, 5>.

Pri vytváraní automatu z ESN sme zobrali každý 50. stav rezervoára ako reakcie na vstupnú trénovaciu postupnosť a z takto získaných 58056 stavov sme technikou K-means vytvorili 10000 klastrov. Vznikol nám teda automat s 10000 stavmi.

Samotné testovanie prebiehalo vybraním prvých 1000 symbolov z testovacej množiny (kniha Genesis) a s pravdepodobnosťou 0,1 sme každý z týchto symbolov

poškodili, t. j. zmenili na iný symbol. Sledovali sme, koľko z poškodených symbolov sa pomocou Viterbiho algoritmu podarilo opraviť, ale takisto aj koľko nakoniec zostalo poškodených symbolov, nakoľko tento algoritmus môže „opraviť“ aj správne symboly. To je spôsobené tým, že máme len model „procesu“, ktorý vygeneroval text biblie a je teda možné, že nejaká iná sekvencia sa javí pravdepodobnejšia ako tá správna.

Experiment sme zopakovali 100 krát – t. j. poškodili sme vždy iných cca 10% z 1000 symbolov a vypočítali priemery. Markovovskému modelu s premenlivou dĺžkou sa z v priemere 100,12 poškodených symbolov podarilo opraviť 81,16, čo predstavuje **81,1%**, a stále poškodených symbolov zostalo **29,38**. Automatu vytvoreného z ESN sa z v priemere 99,84 poškodených symbolov podarilo opraviť 84,32, čo predstavuje **84,5%**, a stále poškodených zostalo **23,6** symbolov.

Tabuľka 2. Príklad opravy textu – hrubým a podčiarknutým sú vyznačené chyby

Pôvodný text	...gathered together unto one place and let the dry land appear and it was so and god called the dry land earth...
Poškodený text	...ga**d**he**t**ed**gg**ogeth**p**r un**n**o **j**ne p_ace and **e**et **j**he dr**f** land appea**m** and it was so **n**nd god _alled the dry land earth...
Opravený text	...gathered together unto **th**e place and let the d**a**y land a_pea**c** and it was so and god called the d**a**y land earth...

V tabuľke 2 uvádzame príklad opravy textu pomocou ESN. Je to úryvok z opravovaných 1000 symbolov z knihy Genesis slúžiacej ako testovacia množina. Dá sa tu všimnúť jednak úspešná oprava silne poškodeného textu (napr. prvé dve slová), ale aj neúspešná oprava (napr. slovo appear), a tiež pokazenie správneho slova (dry).

9 Zhodnotenie

V tejto práci sme na úlohe korekcie poškodeného textu porovnali dva odlišné prístupy – Markovovský model s premenlivou dĺžkou a neurónové siete s echo stavmi. Lepší výkon dosiahli neurónové siete s echo stavmi, najmä z hľadiska menšieho počtu poškodených symbolov v opravenom texte. Hlavnou nevýhodou použitia ESN na túto úlohu je výpočtová náročnosť kroku kvantizácie. Je možné k tomuto pristupovať použitím efektívnejších metód vektorovej kvantizácie ako je K-means, alebo upraviť myšlienku Viterbiho algoritmu tak, aby nebolo potrebné explicitne vytvárať pravdepodobnostný automat. Treba však podotknúť, že po vytvorení pravdepodobnostného automatu z ESN už samotná korekcia textu trvá kratšie ako s automatom vytvoreným z PST vzhľadom na menší počet stavov automatu. Spolu s lepšími dosiahnutými výsledkami je preto používanie korekcie textu pomocou ESN výhodnejšie ako s Markovovským modelom s premenlivou dĺžkou.

Poďakovanie: Tento príspevok vznikol za podpory grantovej agentúry VEGA SR v rámci grantovej úlohy 1/4053/07 a za podpory grantovej agentúry APVT v rámci grantovej úlohy APVT-20-002504.

Literatúra

[1] Jaeger, H.: *Augmented Echo State Networks in a time series prediction task*. In: Advances in Neural Information Processing Systems, MIT-Press, 2002.

[2] Jaeger, H.: *Short Term Memory in Echo State Networks*. Technical report 152, GMD – Forschungszentrum Informationstechnik GmbH, 2002.

[3] Jaeger, H.: *The „echo state" approach to analysing and training recurrent neural networks*. Technical report 148, GMD – German National Research Institute for Computer Science, 2001.

[4] Jaeger, H., Haas, H.: Harnessing nonlinearity: predicting chaotic systems and saving energy in wireless telecommunication. *Science* **304** (2004) 78-80.

[5] Machler, M., Buhlmann, P.: Variable Length Markov Chains: Methodology, Computing and Software. *Journal of Computational and Graphical Statistics* **13** (2004) 435–455.

[6] Natschläger, T., Maass, W., Markram, H.: The "Liquid Computer": A Novel Strategy for Real-Time Computing on Time Series. In: *Special Issue on Foundations of Information Processing of TELEMATIK* **8**(1) (2002) 39-43.

[7] Ron, D., Singer, Y., Tishby, N.: The Power of Amnesia. *Machine Learning* **25** (1996) 117-149.

[8] Tabus, I.: *Recursive Least Squares Estimation. Lecture notes*. www.cs.tut.fi/~tabus/ course/ASP/LectureNew10.pdf

[9] Tiňo, P., Čerňanský, M., Beňušková, Ľ.: Markovian architectural bias of recurrent neural networks. *IEEE Transactions on Neural Networks* **15**(1) (2004) 6-15.

[10] Tiňo, P., Vojtek, V.: Extracting Stochastic Machines from Recurrent Neural Networks. *IEEE Transactions on Neural Networks* **15**(1) (2004) 6-15.

[11] Viterbi, A. J.: Error Bounds for Convolutional Codes and an Asymptotically Optimal Decoding Algorithm. *IEEE Transactions on Information Theory* **13** (1967) 260-269.

Život ve světě symbolů: počítání pomocí membrán

Luděk Cienciala, Alica Kelemenová[1]

Abstrakt. Cílem příspěvku je prezentovat jeden z nových trendů teoretické informatiky, výpočetní model inspirovaný biochemií buněčné struktury nazývaný P systém. Základní úlohou modelu je popsat změny vyvolané biochemickými procesy probíhajícími v hierarchicky organizované membránové struktuře buněk pomocí pravidel umožňujících jejich vyjádření a studium na bázi algoritmických procesů.

1 Úvod

Membránové systémy, které jsou předmětem této studie vycházejí z odpovídající biologické motivace. Shrneme proto úvodem ty biologické fakty, které inspirovaly zavedení modelu a našly v něm svůj odraz.

Život, tak jak ho známe na Zemi, je přímo spojený s pojmem buňky. Buňka je rozměrem velmi malá, ale svým složením a funkcí naprosto jedinečná část živé hmoty, dokonalá svými schopnostmi metabolismu, rozmnožování se a komunikace s okolními buňkami i prostředím, ve kterém žije. Hranice všech organel i samotných buněk tvoří membrány. Membrány určují prostředí, ve kterém se odehrávají biochemické procesy a umožňují více méně kontrolovanou výměnu látek s okolím. Membrány mají schopnost řídit výměnu látek s okolím, ať už na základě pouhé velikosti vlastních molekul, nebo pomocí principu aktivního transportu přes proteinové brány, které nejen vybírají vhodné molekuly, ale propouští je ve směru větší koncentrace látek dovnitř nebo ven. Samotná buňka je schopna řídit svůj život na základě prostředí, ve kterém žije. V jejím spojení s ostatními buňkami v koloniích, orgánech, organismech se skrývá schopnost přizpůsobit se a schopnost komunikovat jak s okolními buňkami, tak s vnějším prostředím. Membrány tvoří hranici oblasti, která obsahuje různé částice (ionty, molekuly, polymery, …) plovoucí ve vodném roztoku bez jakékoliv známky uspořádanosti. Jaderná membrána u eukaryotických buněk uchovává genetický materiál buňky. Ten je skrze stěnu propouštěn pouze k ribozómům, ve kterých je z části DNA vytvořena RNA, jež pak slouží k tvorbě bílkovin a tím i dalších buněčných organel, dalších buněk…

V této práci prezentovaný model reflektuje vztah života a buněk, který u S. Marcuse v [24] nalézáme symbolicky zachycen ve tvaru:

Life = DNA software + membrane hardware.

V roce 1998 zavedl Gheorghe Păun, pod názvem *membránový systém* nový výpočetní model inspirovaný biochemií buněčné struktury. S tímto modelem, časopisecky prvně publikovaným v práci [30], se po roce 2000 střetáme už výlučně pod názvem *P systém*, na počest autora, jednoho z nejvýznamnějších a nejplodnějších autorů současné teoretické informatiky.

[1]Ústav informatiky, Slezská univerzita, Bezručovo nám. 13, 746 01 Opava,
E-mail: {ludek.cienciala,alica.kelemenova}@fpf.slu.cz

P systémy jsou membránové struktury skládající se z membrán hierarchicky vnořených do hlavní membrány. Každá membrána ohraničuje určitou oblast, která může obsahovat další membrány a také objekty, z nichž některé mohou být i stejné. Objekty uvnitř membrán mohou podléhat změnám na základě pravidel přiřazených jednotlivým membránám. Podle těchto pravidel se mění obsah oblastí a tyto změny určují výpočet. Pravidla aplikujeme na objekty (strukturované ve tvaru řetězců nebo nestrukturované) a mohou vyjadřovat různé typy změn objektů jako ku příkladu nahrazení, překřížení, mutace, štěpení, ... Následně se může měnit i pozice nově vytvořených objektů. Objekty mohou setrvat v původní membráně, nebo se mohou posunout do membrány hierarchicky uvnitř nebo vně původní membrány. Na základě zmíněných faktorů byly od roku 1998 zavedeny různé varianty P systémů v závislosti také na specifikaci rozsahu paralelismu typického pro tyto systémy a způsobu získání výsledku výpočtu. Činnost P systému probíhá na bázi stanovených pravidel, v diskrétních krocích, obecně nedeterministicky a paralelně pro všechny objekty membránové struktury. Výpočet začíná počáteční specifikací obsahu jednotlivých membrán P systému, tedy objektů které náleží těmto membránám. Výsledkem výpočtu bývá zpravidla obsah předem specifikované výstupní membrány a to buď všechen, nebo tvořený speciálními výstupními symboly. V některých případech je určující jenom počet těchto objektů.

Výpočet v P systémech nese typické charakteristiky generativního formalismu přepisu v gramatikách (systémech gramatik).

Jiným tradičním modelem výpočetního zařízení jsou akceptující modely automatů nebo strojů. Tyto inspirovaly úvahy Erzsébet Csuhaj-Varjú a György Vaszila [13] a přivedly je k modelu akceptujícího membránového zařízení známého pod názvem P automat. Na rozdíl od jiných P systémů, kde se objekty pomocí pravidel mění (vyvíjí se), jsou P automaty čistě komunikačními výpočetními systémy. Model povoluje pouze přesouvat za určitých podmínek objekty mezi membránami. Pouze hlavní membrána, obal, může propouštět objekty z okolí systému a to je jediný způsob, jak do systému vstupují nové objekty. Výpočet začíná ve vstupním stavu P automatu, který je určen vstupními obsahy jednotlivých oblastí, a končí, když se P automat dostane do koncového stavu, ve kterém oblasti obsahují předepsané (multi)množiny objektů. Výsledkem výpočtu je posloupnost objektů, které systém přijal v průběhu úspěšného výpočtu, ze vstupního stavu do stavu koncového. P automat tedy pracuje podobně jako tradiční automat: posloupnost vstupujících objektů v P automatu odpovídá řetězci čtenému ze vstupní pásky konečného automatu a oblasti s objekty v P automatu odpovídají stavům konečného automatu.

P systémy je možno chápat jako gramatické systémy [11] komunikující prostřednictvím specifické stromové struktury komponent a se specifickou sadou pravidel v komponentách. Proto bylo podnětným inspirovat se jedním z trendů vývoje v gramatických systémech, který sleduje schopnost co nejjednodušších komponent umístěných ve společném prostředí a komunikujících navzájem výlučně prostřednictvím prostředí provádět výpočty mnohem náročnější než ty, které jsou schopny provést samotné komponenty. Takové skupiny komponent nazýváme koloniemi. Adaptace úvah o koloniích v membránového prostředí vedla k P koloniím v [22], modelu, který v „P světě“ reflektuje požadavek komponent na jednoduchost a jejich umístěním ve společném prostředí zvětšuje výpočetní sílu skupiny komponent.

P kolonie byly zavedeny v práci [22] v roce 2004. Daný model zachycuje činnost kolonie „jednomembránových organizmů.“ V jejich společném prostředí se nacházejí spe-

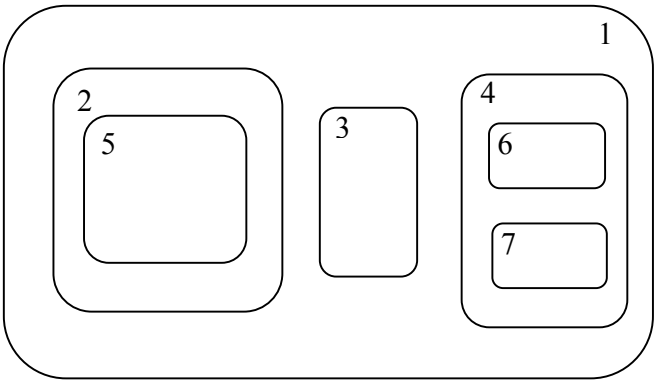
ciální objekty, které budeme značit *e*. V případě, že si výpočet symbol *e* z prostředí vyžádá, je takovýto symbol vždy k dispozici.

V prostředí, které v P koloniích slouží jako komunikační kanál i „skladiště“ objektů, se nacházejí výlučně jednomembránové systémy, nazývané agenty. Počet objektů uvnitř každého agenta je stejný pro všechny agenty kolonie a v průběhu výpočtu se nemění. Činnost agenta je řízena vlastní množinou programů, tvořených přepisujícími, komunikačními a/nebo kontrolními pravidly. Počet pravidel programu se shoduje s počtem objektů uvnitř agenta. V každém diskrétním okamžiku, kdy se realizuje výpočetní akce jsou všechny objekty uvnitř agenta změněny nebo přesunuty. Analogicky jako v P systémech obecně i u stávajícího modelu můžeme specifikovat různé stupně paralelizmu, zkoumat možnosti P kolonií ve vztahu k počtu objektů v agentech i podle skladby pravidel v programech, případně můžeme vyčlenit některé objekty jako koncové.

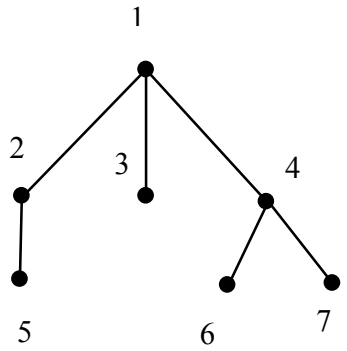
V následujících částech práce představíme formalizmus některých typů P systémů, ilustrujeme je na příkladech a shrneme některé známé poznatky do vět. Čtenáře, který se zajímá o další detaily, odkážeme na citovanou literaturu. Stručné informaci o stávajícím výzkumu P automatů a P kolonií budou věnovány poslední dvě kapitoly práce, opět doplněny odkazy na literaturu v této oblasti.

2 Membránové struktury

Membránová struktura je složená z buněčných membrán hierarchicky vnořených do hlavní membrány, kterou nazýváme obal.



Obr. 1. Znázornění membránové struktury pomocí Vennových diagramů



Obr. 2. Znázornění membránové struktury pomocí grafu

Membránovou strukturu zapisujeme dvojicemi závorek označených číslem membrány tak, že každý pár závorek vyjadřuje umístění membrány v hierarchicky uspořádaném membránovém systému. Graficky je možno znázornit membránovou strukturu pomocí Vennových diagramů, ve kterých je každá dvojice membrán disjunktní nebo je jedna podmnožinou druhé, nebo pomocí stromu s kořenem odpovídajícím obalu a listy reprezentujícími elementární membrány, tedy membrány, které neobsahují žádné další membrány. Závorkový výraz $[_1 [_2 [_5]_5]_2 [_3]_3 \ [_4 [_6]_6 [_7 \ ]_7]_4]_1$ reprezentuje membránovou strukturu znázorněnou Vennovým diagramem na obr. 1 a pomocí stromu na obr. 2.

3 Membránové systémy - neformální popis

Membránový systém je tvořen membránovou strukturou, ve které slouží jednotlivé membrány na umístění objektů. Objekty dané oblasti jsou v modelech reprezentovány prvky konečné množiny, nebo eventuálně jako řetězce náležící do konečné množiny. Vyvíjejí se podle daných evolučních pravidel, přiřazených k jednotlivým oblastem. To znamená, že daná pravidla jsou aplikována pouze na objekty v dané oblasti. Pravidla obsahují také cílovou indikaci specifikující membránu/oblast, do které bude objekt vytvořený po aplikaci pravidla náležet. Důležitou vlastností vývoje membránového systému je, že použití evolučních pravidel vzhledem k jednotlivým objektům probíhá paralelně a jejich aplikace je regulována danými relacemi priority. Objekty se mohou vyvíjet nezávisle nebo ve spolupráci s ostatními objekty. Speciální třídu objektů tvoří katalyzátory, kterých přítomnost umožňuje vývoj jiných objektů, sami však změně nepodléhají. Objekty mohou přecházet přes membrány v obou směrech. Mohou přejít přes membránu, která ohraničuje oblast, ve které se nacházejí, směrem ven anebo mohou přejít přes membránu, která se nachází uvnitř dané oblasti. Membrány se můžou také rozpustit. Potom se všechny objekty, které byly v rozpuštěné membráně stanou objekty membrány umístěné v hierarchii bezprostředně nad původní membránou. Současně evoluční pravidla rozpuštěné membrány zanikají. Obal systému se nemůže rozpustit.

Vývoj membránového systému charakterizujeme zápisem konečné posloupnosti aktuálních obsahů jednotlivých membrán systému. Tyto údaje o stavu systému určující jeho další činnost nazýváme konfigurací systému.

Počáteční konfigurace a způsob používání evolučních pravidel určují průběh vývoje a tedy i výpočet realizovaný membránovým systémem. Výpočet považujeme za ukončený tehdy, když neexistuje pravidlo použitelné pro další výpočet. Výstup výpočtu může být určen výstupní membránou nebo prostředím. V prvním případě je výstupem číslo určující počet všech objektů nacházejících se ve výstupní membráně v cílové konfiguraci. Ve druhém případě je výsledkem množina řetězců, které získáme ze symbolů opouštějících systém tj. obal, v přirozeném pořadí. Pokud symboly opouštějí systém současně, pak je v řetězci akceptována každá jejich permutace.

Po tomto neformálním popisu membránových systému přistoupíme v dalších kapitolách k jejich formalizaci ve tvaru P systémů. Úvodem budeme pracovat se systémy s jednoduchými, vnitřně nestrukturovanými objekty reprezentovanými symboly. Budou předmětem kapitoly 4. V kapitole 5 se budeme zabývat objekty se strukturou ve tvaru řetězců.

4 P systémy s objekty ve tvaru symbolů

Definice 1: P systém stupně m, $m \geq 1$ je struktura

$$\Pi = (V, T, C, \mu, M_1, ..., M_m, (R_1, \rho_1), ..., (R_m, \rho_m), i_0),$$ kde

- V je konečná abeceda objektů.
- $T \subseteq V$ je výstupní abeceda.
- $C \subseteq V$, $C \cap T = \emptyset$, je množina katalyzátorů.
- μ je membránová struktura sestávající z m oblastí, číslovaných od 1 do m, specifikovaných membránami.
- M_i, $1 \leq i \leq m$, je multimnožina nad V přiřazená i-té oblasti struktury μ.

- R_i, $1 \le i \le m$, je konečná množina evolučních pravidel nad V, přiřazená i-té oblasti membránové struktury μ.
  Evoluční pravidlo je dvojice (u,v), kterou zapisujeme ve tvaru $u \to v$, kde u je řetězec nad V, v je řetězec nad $V \times (\{zde, ven\} \cup \{do_j / 1 \le j \le m\})$ ukončený případně speciálním symbolem $\delta \notin V$, reprezentujícím rozpuštění membrány.
- ρi, $1 \le i \le m$, *je relace nad Ri, která specifikuje priority mezi pravidly z množiny Ri.*
- i_0 je cílová membrána, $1 \le i_0 \le m$ nebo vnější oblast struktury reprezentována hodnotou $i_0 = \infty$.

Množinu M_i můžeme zapsat také jako řetězec w_i objektů, které se v M_i nacházejí. Při této reprezentaci nezáleží na pořadí symbolů ve slově w_i.

Příklad 1. Výpočet kvadratické funkce.

Nechť $\Pi_1 = (V, T, C, \mu, w_1, w_2, w_3, w_4, (R_1, \rho_1), (R_2, \rho_2), (R_3, \rho_3), (R_4, \rho_4), 4)$ [2] *je P systém, kde*

$V = \{a, b, b_1, b_2, c, f\}$, $T = \{c\}$, $C = \emptyset$,

$\mu = [_1\ [_2\ [_3\ ]_3\ [_4\ ]_4\ ]_2\ ]_1$,

$w_1 = \varepsilon$, $R_1 = \emptyset$, $\rho_1 = \emptyset$,

$w_2 = \varepsilon$,

$R_2 = \{b_1 \to b_2,\ b_2 \to b,$

$b \to b(c, do_4),$

$r_1: ff \to f,\ r_2: f \to \delta\}$

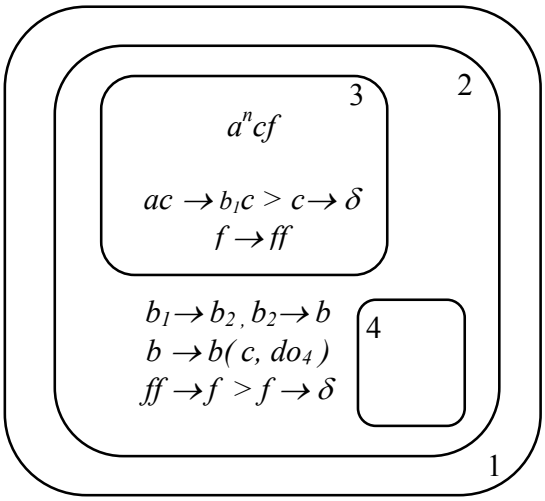
$\rho_2 = \{r_1 > r_2\}$

$w_3 = a^n cf$,

$R_3 = \{r_3:\ ac \to b_1 c,\ r_4: c \to \delta,$

$f \to ff\},\ \rho_3 = \{r_3 > r_4\}$

$w_4 = \varepsilon$, $R_4 = \emptyset$, $\rho_4 = \emptyset$.



Obr. 3. P systém Π_1

Membránová struktura μ a multimnožiny $w_1, \ldots, w_m$ určují počáteční konfiguraci $(\mu, w_1, \ldots, w_m)$ P systému Π_1. Membrány zachovávají v průběhu celého výpočtu své označení. Obecně, konfigurace $(\mu', w'_{i_1}, \ldots, w'_{i_k})$ je určena membránovou strukturou μ' která vznikne z μ odstraněním všech membrán různých od $i_1, \ldots, i_k$. P systém přechází z jedné konfigurace do druhé paralelním použitím evolučních pravidel.

Pro konfigurace $c_1 = (\mu', w'_{i_1}, \ldots, w'_{i_k})$ a $c_2 = (\mu'', w''_{j1_1}, \ldots, w''_{j_l})$ membránového systému Π píšeme $c_1 \Rightarrow c_2$, jestliže použitím pravidel přechází systém z konfigurace c_1 do

[2] Při zápisu pravidel budou dvojice tvaru *(a, zde)* nahrazeny první složkou, písmenem *a*. Polohu „zde“ budeme vynechávat.

konfigurace c_2. Použití vývojových pravidel je paralelní. To znamená, že všechny objekty ze všech membrán, které mohou být subjekty lokálních vývojových pravidel, se vyvíjejí současně. Na přepisu se podílí pravidlo $u \to v$ z R_i s největší prioritou. Použití tohoto pravidla v oblasti i znamená odstranění všech objektů určených slovem u z této oblasti a přidání objektů určených prvními komponenty slova v do oblastí určených druhými komponenty tohoto slova. Pokud v obsahuje

- (a, zde), pak a umístíme do stejné oblasti, ve které bylo u;
- (a, ven), pak a umístíme do oblasti, která je bezprostředně v hierarchii nad oblastí, ve které bylo u;
- (a, do_i), pak a dáme do oblasti, která je označená i.

Pokud se v pravidle vyskytuje speciální symbol δ, pak se po aplikaci pravidla membrána rozpustí.

Výpočet je úspěšný, jestliže se P systém zastaví, tj. neexistuje pravidlo, které by bylo možné aplikovat na cílovou konfiguraci. Výstup P systému je určen počtem objektů výstupní abecedy ve výstupní membráně cílové konfigurace.

Příklad 1 (pokračování): Výpočet kvadratické funkce

Počáteční konfigurace P systému je $(\varepsilon, \varepsilon, a^n cf, \varepsilon)$.

P systém začínající v této konfiguraci ukončí výpočet s n^2 objekty c ve výstupní (tj. čtvrté) membráně. Ve vztahu k počtu symbolů v počáteční konfiguraci realizuje výpočet kvadratické funkce.

Průběh výpočtu:

V prvním kroku je možné aplikovat jenom pravidla ve třetí oblasti. Vzhledem na prioritu pravidel použijeme pravidla $ac \to b_1c$, $f \to ff$ na přepsání jednoho výskytu písmen a, c, f. Situace se opakuje pokud se ve třetí oblasti nachází objekt a.

$$(\varepsilon, \varepsilon, a^n cf, \varepsilon) \Rightarrow (\varepsilon, \varepsilon, a^{n-1} b_1 cff, \varepsilon) \Rightarrow (\varepsilon, \varepsilon, a^{n-2} b_1 b_1 cffff, \varepsilon) \Rightarrow^{n-2} (\varepsilon, \varepsilon, b_1^n c f^{2^n}, \varepsilon) \Rightarrow$$

V následujícím kroku použitím pravidla $c \to \delta$ dojde k rozpuštění třetí membrány a objekty $b_1^n f^{2^{n+1}}$ přejdou do oblasti 2. Fakt, že membrána 3 po tomto kroku výpočtu v struktuře není přítomna bude označen v konfiguraci symbolem - .

$$\Rightarrow (\varepsilon, b_1^n f^{2^{n+1}}, -, \varepsilon) \Rightarrow$$

V dalším kroku se přepíšou všechna b_1 na b_2 po aplikaci pravidla $b_1 \to b_2$ a počet objektů f se sníží použitím pravidla $ff \to f$ na polovinu, tzn. f^{2^n}.

$$\Rightarrow (\varepsilon, b_2^n f^{2^n}, -, \varepsilon) \Rightarrow (\varepsilon, b^n f^{u(n-1)}, -, \varepsilon) \Rightarrow$$

kde $u(k) = 2^k$.

Použití pravidla $b \to b(c, do_4)$, bude mít za následek, že se všechna b ve druhé oblasti zachovají a za každé b bude vloženo c do oblasti výstupní membrány 4. Oblast 4 bude nyní obsahovat c^n a 2^{n-2} symbolů f bude v oblasti 2.

$$\Rightarrow (\varepsilon, b^n f^{u(n-2)}, -, c^n) \Rightarrow$$

V každém dalším kroku se do oblasti 4 přidává n symbolů c současně se snižováním počtu f pomocí pravidla $ff \to f$.

$$\Rightarrow^* (\varepsilon, b^n f, -, c^{n(n-1)}) \Rightarrow^* (b^n, -, -\, c^{n^2})$$

V posledním kroku se použijí pravidla $b \rightarrow b\ (c, do_4)$ a $f \rightarrow \delta$, které rozpustí membránu 2. Výpočet je ukončen, protože oblastem 1 a 4, které jediné zbyly, nejsou přiřazeny žádné přepisovací pravidla. Ve výstupní membráně se nachází n^2 objektů c.

Třídu množin čísel generovaných P systémy stupně n se symboly označíme $TP_n\ (Coo)$, $TP_n\ (Cat)$, $TP_n\ (nCoo)$, kde Coo je pro systémy s interaktivními pravidly, Cat pro systémy používající interaktivní pravidla pouze v roli katalyzátoru a $nCoo$ pro systémy bez interakce. Pokud umožníme rozpouštění membrán, vyznačíme tuto skutečnost v závorce jako δ ve druhé komponentě.

Věta [34]: $TP_2\ (Coo) = TP_2\ (Cat) = RE.$
$TP_1\ (nCoo) \subset TP_2\ (nCoo) \subset \ldots\ T\ (nCoo) \subset T\ (nCoo,\delta) \subset RE$
$TP_1\ (nCoo,\delta) \subset TP_2\ (nCoo,\delta) \subset \ldots\ \subset TP(nCoo,\delta)$
$TP_1\ (nCoo) \subset TP_1\ (Cat) \subset TP_1\ (Coo,\delta) \subset RE.$

5 P systémy s objekty ve tvaru řetězce

V modelu uvažovaném v předchozí kapitole jsme objekty v membránách považovali za atomické, bez vnitřní struktury a byly reprezentovány písmenem. Model uvažovaný ve stávající kapitole bude podrobnější, s objekty které budou mít strukturu ve tvaru řetězce. Pro manipulaci s řetězci v těchto systémech využijeme různé typy pravidel, které známe z teorie formálních jazyků, z DNA výpočtů, atd. Jsou to například operace přepisování/substituce, sestřižení, vkládaní, odstranění. V následujících částech kapitoly se budeme zabývat P systémy pracujícími s objekty ve tvaru řetězců, které se vyvíjejí na bázi zmíněných pravidel. Vývoj jednotlivých součástí P systémů může být jak sekvenční tak i paralelní. V modelech se může vyskytnout paralelizmus na třech úrovních: na úrovni přepisování symbolů v řetězci, na úrovni řetězců každé z oblastí a na úrovni oblastí membránového systému.

Výstup v těchto modelech může být určen obsahem cílové membrány po ukončení výpočtu nebo řetězci, které opustily systém. Může to být také počet řetězců v každém z předchozích případů. V následujících podkapitolách obeznámíme čtenáře s třemi základními typy P systémů s objekty ve tvaru řetězce. Budou to přepisující P systémy, SP systémy (systémy se s operací sestřižení) a systémy s dělícími se objekty.

Pro *přepisující systémy* je typický bezkontextový tvar pravidel, sekvenční přepisování symbolů v řetězcích, paralelní přepisování řetězců v dané oblasti, paralelní přepisování oblastí. Výstup je tvořen řetězci, které opouštějí systém.

SP systémy pracují s pravidly sestřihu, které aplikují na přepis konkrétního řetězce sekvenčně, krok výpočtu probíhá paralelně pro všechny řetězce v oblasti, které mohou být přepsány a také současně ve všech oblastech systému. Výstupy tvoří řetězce, které opouštějí systém.

Přepis v *P systémech s dělícími se objekty* určují pravidla replikace, štěpení, mutace a rekombinace, která jsou aplikována na konkrétní řetězec sekvenčně, krok výpočtu probíhá paralelně pro všechny řetězce v oblasti, které mohou být přepsány a také současně ve všech oblastech systému. Výsledek výpočtu je umístěn v cílové membráně, případně ve vnějším prostředí systému.

5.1 Přepisující P systémy

Přepisující P systémy pracují tak, že mění řetězce na základě pravidel, které umožňují substituci symbolu v řetězci novým řetězcem a určují také membránu, v které má být vytvořený objekt umístěn. Pravidlo systému má tvar $A \rightarrow (\alpha, X)$, kde A je neterminální symbol, α je řetězec složený ze symbolů abecedy V, $X \in \{ do, ven, zde \}$.

Definice 2: Přepisovací P systém je struktura

$$\Pi = (V, T, \mu, L_1, \ldots , L_m, (R_1, \rho_1), \ldots, (R_m, \rho_m)),$$ kde:

- V je konečná abeceda.
- $T \subseteq V$ je terminální abeceda.
- μ je membránová struktura s m membránami označenými $1, 2, \ldots, m$.
- L_i, $1 \leq i \leq m$, je konečný jazyk nad V (počáteční řetězec v oblasti i).
- R_i, $1 \leq i \leq m$, je konečná množina bezkontextových pravidel umístěná v oblasti určené membránou i.
- ρ_i, $1 \leq i \leq m$, je relace částečného uspořádání na množině pravidel R_i.

Vzhledem k tomu, že konečné množiny řetězců umístěné v jednotlivých oblastech jsou určující pro další vývoj systému, definujeme konfiguraci přepisovacího P systému jako $(m+1)$ tici těchto množin řetězců, korespondujících s oblastmi membránové struktury, přičemž poslední $(m+1)$*-ní* komponenta odpovídá vnějšku membránového systému a obsahuje řetězce vycházející ven z membránové struktury.

Počáteční konfigurace má tedy tvar $c_0 = (L_1, \ldots, L_m, \emptyset)$.

Pro dvě konfigurace $C_1 = (L_1, \ldots, L_m, L_{m+1})$ a $C_2 = (L_1', \ldots, L_m', L_{m+1}')$ systému Π píšeme $C_1 \Rightarrow C_2$, jestliže použitím pravidel přechází systém z konfigurace c_1 do konfigurace c_2. Použití pravidel v jednom kroku výpočtu je dáno takto:

- na úrovni přepisu jednotlivých slov je sekvenční, tj. jedno pravidlo z oblasti výskytu slova v něm přepíše jeden výskyt symbolu určeného levou stranou pravidla,
- na úrovni přepisu slov v oblasti je paralelní, v jednom kroku se přepisují všechna slova nacházející se ve společné oblasti,
- na úrovni přepisu oblastí paralelně.

Jazyk určený přepisujícím P systémem obsahuje ty řetězce, které opustily membránovou strukturu v průběhu výpočtu.

Příklad 2. [23] P systém generující jazyk $L = \{ a^n b^n c^n \mid n \geq 1 \}$.

Mějme přepisující P systém

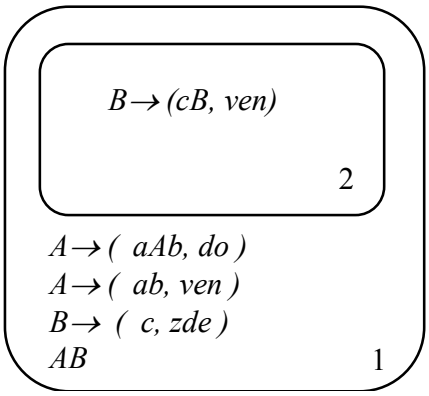
$\Pi_2 = (V, T, \mu, L_1, L_2, (R_1, \emptyset), (R_2, \emptyset))$, kde

$V = \{ A, B, a, b, c \}$, $T = \{ a, b, c \}$, $\mu = [_1 \; [_2 \;]_2 \;]_1$,

$L_1 = \{ AB \}$, $L_2 = \emptyset$,

$R_1 = \{ A \rightarrow (aAb, do), A \rightarrow (ab, ven),$
$B \rightarrow (c, zde) \}$,

$R_2 = \{ B \rightarrow (cB, ven) \}$.



Obr. 4. P systém Π_2

Počáteční konfigurace je určena trojicí $(AB, \emptyset, \emptyset)$.

V *1.* kroku může být použito například pravidlo $A \rightarrow (aAb, do)$, které do oblasti 2 přesune řetězec

aAbB. Ve *2*. kroku je v oblasti 2 použito jediné pravidlo $B \rightarrow (cB, ven)$, které přesune vzniklý řetězec *aAbcB* zpět do oblasti 1.

Formálně

$$(AB, \emptyset, \emptyset) \Rightarrow (\emptyset, aAbB, \emptyset) \Rightarrow (aAbcB, \emptyset, \emptyset) \Rightarrow$$

Tímto způsobem může výpočet pokračovat. Po *2n* krocích je v oblasti *1* řetězec $a^nAb^nc^nB$. Použitím pravidla $B \rightarrow (c, zde)$ v oblasti *1*, se v této oblasti vytvoří řetězec $a^nAb^nc^{n+1}$. Po aplikaci pravidla $A \rightarrow (ab, ven)$ vzniká řetězec $a^{n+1}b^{n+1}c^{n+1}$, který opouští systém, je tedy výstupem výpočtu.

$$\Rightarrow * (a^nAb^nc^nB, \emptyset, \emptyset) \Rightarrow (a^nAb^nc^{n+1}, \emptyset, \emptyset) \Rightarrow (\emptyset, \emptyset, a^{n+1}b^{n+1}c^{n+1}).$$

$L(\Pi_2)=\{a^nb^nc^n : n \geq 1\}$.

Třídu jazyků generovaných přepisujícími P systémy stupně *m* používající prioritu mezi pravidly označíme $RP_m(Pri)$, pokud mezi pravidly neuvažujeme o prioritě budeme používat označení $RP_m(nPri)$.

***Věta* [30, 34]:** $RP_2(Pri) = RE$.

$RP_1(nPri) \subset RP_1(Pri) \subset RE$

$RP_1(nPri) \subset RP_2(nPri) \subset RP_3(nPri) \ldots \subset RP(nPri) \subset RE$

5.2 SP systémy

Vývoj řetězců se u SP systémů uskutečňuje na základě pravidla známého z oblasti DNA výpočtů, pravidla sestřihu. Pravidlo sestřihu je určeno řetězcem $r = u_1\#u_2\$u_3\#u_4$ obsahujícím dva speciální symboly #, $, které v řetězci *r* oddělují čtyři podřetězce u_1, u_2, u_3, u_4. Operace sestřihu je definována pro dvojice řetězců (x, y) a jejím výsledkem jsou opět dvojice řetězců (w, z) vytvořené tak, že se rozstřihne řetězec *x* mezi u_1 a u_2 a řetězec *y* mezi u_3 a u_4 a tyto části vytvoří dva nové řetězce spojením vzájemně zaměněných prvních a druhých částí původních slov. Formálně:

Pro pravidlo sestřihu $r = u_1\#u_2\$u_3\#u_4$, kde $u_1, u_2, u_3, u_4 \in V^+$, a dvojici slov $x, y \in V^+$ platí

$$(x, y) \vdash_r (w, z) \text{ právě tehdy když } x = v_1u_1u_2v_2,\ y = v_3u_3u_4v_4,$$
$$w = v_1u_1u_4v_4,\ z = v_3u_3u_2v_2$$

Zavedeme P systémy, ve kterých ke změně řetězců dochází na základě pravidel sestřihu. Budeme je označovat jako SP systémy.

Definice 3: SP systém je struktura

$$\Pi = (V, T, \mu, L_1, \ldots, L_m, R_1, \ldots, R_m), \text{ kde}$$

- V je konečná abeceda.
- T je výstupní abeceda, $T \subseteq V$.
- μ je membránová struktura z *m* membrán $1, \ldots, m$.
- L_i, $1 \leq i \leq m$ jsou konečné jazyky nad abecedou V.
- R_i, $1 \leq i \leq m$ jsou konečné množiny evolučních pravidel asociovaných s oblastmi
- $1, .., m$ z μ ve tvaru $(r; c_1, c_2)$, kde $r = u_1\#u_2\$u_3\#u_4$ je pravidlo sestřihu nad V a
- $c_1, c_2 \in \{zde, ven, do\}$.

V SP systémech vycházíme z předpokladu, že pokud řetězec patří do oblasti, pak v dané oblasti je v libovolném počtu kopií.

Stejně jako u přepisujících P systémů, i v případě SP systémů konečné množiny řetězců umístěné v jednotlivých oblastech určují další vývoj systému. Konfigurace SP systému je $(m+1)$-tice těchto množin řetězců, korespondujících s oblastmi membránové struktury, přičemž poslední $(m+1)$-ní komponenta odpovídá vnějšku membránového systému a obsahuje řetězce vycházející ven z membránové struktury.

Krok výpočtu je určený následujícími změnami konfigurací: Jestliže pro $x, y \in L_i$ a $r = (u_1\#u_2\$u_3\#u_4;\ c_1,c_2) \in R_i$ a platí $(x, y) \vdash_r (w, z)$, pak je w umístěn do oblasti c_1 a z do oblasti c_2. Po aplikaci pravidla jsou řetězce x a y v původní oblasti stále k dispozici, ale jestli řetězce w, z jsou poslány ven z oblasti, nejsou v oblasti už žádné jejich kopie. Krok výpočtu probíhá paralelně pro všechny řetězce v oblasti, které mohou být přepsány a také současně ve všech oblastech systému. Přepis konkrétního řetězce je sekvenční, jedno pravidlo je možné použít na každém kroku výpočtu tj. probíhající změna řetězce a je určena (nedeterministicky) jedním pravidlem.

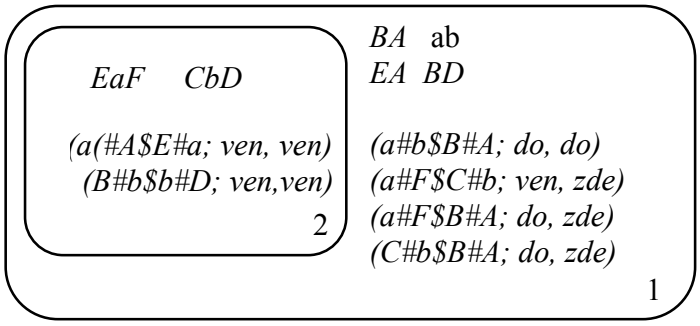
Výstup výpočtu je složen ze všech řetězců nad T, které opustily systém kdykoliv během výpočtu. Označíme $L(\Pi)$ jazyk všech řetězců tohoto typu. V tomto typu P systému nepožadujeme zastavení výpočtu.

Příklad 3. SP systém generující jazyk $L = \{a^n b^n \mid n \geq 2\}$

Mějme SP systém $\Pi_3 = (V, T, \mu, L_1, L_2, R_1, R_2)$, kde

$V=\{ a, b, A, B, C, D, E, F \}$,
$T = \{ a, b \}$,
$\mu =[_1\ [_2\ ]_2\ ]_1$
$R_1 = \{ \ (a\#b\$B\#A;\ do, do),$
$(a\#F\$C\#b;\ ven, zde),$
$(a\#F\$B\#A;\ do, zde),$
$(C\#b\$B\#A;\ do, zde) \}$
$R_2 = \{ (a\#A\$E\#a;\ ven, ven),$
$(B\#b\$b\#D;\ ven, ven) \}$

$L_1=\{ ab, BA, EA, BD \}$,
$L_2 = \{ EaF, CbD \}$,



Obr. 5. SP systém Π_3

Výpočet: Oblast 1 obsahuje řetězce ab, BA a pravidlo $(a\#b\$B\#A;\ do, do)$, proto $(ab, BA) \vdash (aA, Bb)$. Řetězce aA, Bb vytvořeny v tomto kroku budou umístěny v oblasti 2. Oblast 2 v počáteční konfiguraci neobsahuje slova vhodná pro aplikaci pravidel z R_2.

$$(L_{10}, L_{20}, \varnothing) \Rightarrow (L_{11}, L_{21}, \varnothing),$$

kde $L_{11} = L_{10}$, $L_{21} = L_{20} \cup \{ aA, Bb \} = \{ aA, Bb, EaF, CbD \}$.

V dalším kroku výpočtu je ve druhé membráně na řetězce aA, EaF aplikovatelné pravidlo $(a\#A\$E\#a;\ ven, ven)$ a na řetězce Bb, CbD pravidlo $(B\#b\$b\#D;\ ven, ven)$ přičemž platí

$$(aA, EaF) \vdash (aaF, EA) \quad \text{a} \quad (Bb, CbD) \vdash (BD, Cbb).$$

Nově vzniklé řetězce aaF, EA, BD, Cbb budou umístěny do oblasti *1*. V první oblasti probíhají stejné kroky jako v prvním kroku výpočtu. Proto platí

$$(L_{11}, L_{21}, \varnothing) \Rightarrow (L_{12}, L_{22}, \varnothing),$$

kde $L_{12} = L_{11} \cup \{ aaF, EA, BD, Cbb \} = \{ ab, BA, EA, BD, aaF, Cbb, EA, BD \}$,

$L_{22} = L_{21} = \{ aA, Bb, EaF, CbD \}$.
V dalším kroku může být v oblasti 1 aplikováno některé z pravidel

$$(a\#F\$C\#b;\ ven,\ zde),\quad (a\#F\$B\#A;\ do,\ zde),\quad (C\#b\$B\#A;\ do,\ zde)$$

s následujícím účinkem na řetězce z jazyka první komponenty

$$(aaF, Cbb) \vdash (aabb, CF),\quad (aaF, BA) \vdash (aaA, BF),\quad (Cbb, BA) \vdash (CA, Bbb).$$

Vzniká terminální řetězec *aabb*, který opouští systému, resp. řetězce *aaA*, *Bbb* umístěné v druhé oblasti.

$$(L_{12}, L_{22}, \emptyset) \Rightarrow (L_{13}, L_{23}, \{ aabb \}),$$

kde $L_{13} = L_{12} \cup \{ CF, BF, Bbb \} = \{ ab, BA, EA, BD, aaF, Cbb, EA, BD, CF, BF, Bbb \}$,
$L_{23} = L_{22} \cup \{ aA, aaA, Ca, Bb, EaF, CbD \} = \{ aA, aaA, Bb, Bbb, EaF, CbD, Ca \}$.

Tímto způsobem systém vytváří v první oblasti řetězce $a^{n+1}F$ a Cb^{n+1} a v dalším kroku použitím pravidla $(a\#F\$C\#b;\ ven,\ zde)$ produkuje výstupný řetězec $a^{n+1}b^{n+1}$ systému. $L(\Pi_3) = \{ a^n b^n \mid n \geq 2 \}$.
Třídu jazyků generovaných SP systémy stupně *m* označíme $SP_m(i/o)$.

Věta [30, 34]: $SP_2 (i/o) = RE$.

$$SP_1 (nPri, i/o) \subset SP_2 (nPri, i/o) \subset SP_3 (nPri, i/o) \subset SP_4 (nPri, i/o) = RE.$$

5.3 P systémy s dělícími se objekty

V modelech P systémů se kterými jsme se zabývali v předchozích kapitolách jsou výsledky výpočtu určeny následovně: V P systémech, které mají za objekty symboly jsou výsledkem výpočtů přirozená čísla (nebo vektory přirozených čísel). V P systémech, které mají objekty reprezentovány řetězci jsou výsledkem výpočtů (multi)množiny řetězců. Kombinací těchto přístupů, můžeme při práci s multimnožinami řetězců za výsledek výpočtu považovat počet řetězců přítomných ve výstupní membráně po ukončení výpočtu. Tento přístup nacházíme u systémů, které umožňují měnit počet výskytů řetězců v průběhu výpočtu. Uvedeme čtyři pravidla pro práci s řetězci, kterých zavedení vyplynulo z biologické motivace modelu a jejich aplikaci v procesu výpočtu. Jsou to pravidla replikace, štěpení, mutace a rekombinace.

	pravidlo	krok výpočtu
REPLIKACE	$r: a \rightarrow u_1 // u_2$ $a \in V,\ u_1, u_2 \in V^+$	$x_1 a x_2 \Rightarrow_r (x_1 u_1 x_2, x_1 u_2 x_2)$. $x_1, x_2 \in V^*$
ŠTĚPENÍ	$r : a \rightarrow u_1 / u_2$ $a \in V,\ u_1, u_2 \in V^+$	$x_1 a x_2 \Rightarrow_r (x_1 u_1,\ u_2 x_2)$. $x_1, x_2 \in V^*$
MUTACE	$r: a \rightarrow u$ $a \in V,\ u \in V^+$	$x_1 a x_2, \Rightarrow_r x_1 u x_2$. $x_1, x_2 \in V^*$
REKOMBINACE	řetězec překřížení $z \in V^+$	$(x_1 z x_2, y_1 z y_2) \Rightarrow_z (x_1 z y_2,\ y_1 z x_2)$. $x_1, x_2, y_1, y_2 \in V^*$

Aplikace pravidla replikace a štěpení zvyšuje počet řetězců, použití pravidla mutace a rekombinace počet řetězců v P systému zachovává. Počet řetězců může v membránové struktuře také klesat, přechodem řetězců ven ze systému.

Při práci s multimnožinami řetězců s konečnou násobností výskytu určenou $\sigma : V^* \rightarrow N$ budeme jejich násobnost v množině A zapisovat ve tvaru

$$A = \{ (x_1, \sigma(x_1)), ..., (x_k, \sigma(x_k)) \}, \text{ kde } \sigma(x_i) > 0 \text{ pro } x_i \in V^*,\ 1 \le i \le k.$$

Definice 4: P systém stupně m, kde $m \ge 1$, *s dělícími se objekty* je struktura

$$\Pi = (V, \mu, A_1, ..., A_m, (R_1, S_1, M_1, C_1),..., (R_m, S_m, M_m, C_m), i_0), \text{ kde}$$

- V je abeceda.
- μ je membránová struktura z m membrán $1,..., m$.
- $A_1,..., A_m$ jsou konečné multimnožiny nad V^* asociované s oblastmi z μ.
- R_i, S_i, M_i, C_i pro i, $1 \le i \le m$, jsou konečné množiny pravidel replikace, štěpení, mutací a bloků překřížení tvaru:
  - pravidlo replikace: $(a \to u_1 // u_2; c_1, c_2)$, pro $c_1, c_2 \in \{ zde, ven, do \}$;
  - pravidlo štěpení: $(a \to u_1 / u_2; c_1, c_2)$, pro $c_1, c_2 \in \{ zde, ven, do \}$;
  - pravidlo mutace: $(a \to u; c)$, pro $c \in \{ zde, ven, do \}$;
  - blok překřížení: $(z ; c_1, c_2)$, pro $c_1, c_2 \in \{ zde, ven, do \}$.
- $i_0 \in \{ 1, 2, ..., m \}$ je označení výstupní membrány systému.

$(A_1, ..., A_m)$ určuje počáteční konfiguraci systému. Aplikace operací definovaných pomocí pravidel (R_i, S_i, M_i, C_i), $1 \le i \le m$ určuje změnu konfigurace. Řetězec, který vstupuje do operace zaniká. Jeho násobnost se sníží o jedničku. Násobnost řetězců produkovaných operacemi se zvýší. Krok odvození ve vztahu k řetězci je sekvenční. Na přepis každého řetězce použijeme v daném kroku pouze jednu operaci. Vytvořený řetězec se umístí do oblasti, určené použitým pravidlem.

Výsledek výpočtu je určen počtem řetězců v cílové oblasti i_0 po zastavení výpočtu. Výpočet, který se nezastaví, neurčuje žádný výstup.

Symbolem NWP_n označíme množinu čísel vypočitatelnou P systémy s dělícími se objekty.

Příklad 4. [34] Uvedeme příklad určení hamiltonovské cesty grafu pomocí P systémů s dělícími se objekty. Nechť $\gamma = (U, E)$ *je orientovaný graf*, který neobsahuje cykly (i ,i), $i \in U$. Předpokládejme, že U obsahuje n vrcholů, označených čísly $1, 2, ..., n$, vrcholy 1 a n nesousedí a maximální počet hran vycházejících z uzlu je roven $n - 1$. Sestrojíme P systém s dělícími se objekty, používající výhradně replikace a mutace. V zápisu systému vynecháváme prázdné množiny pravidel

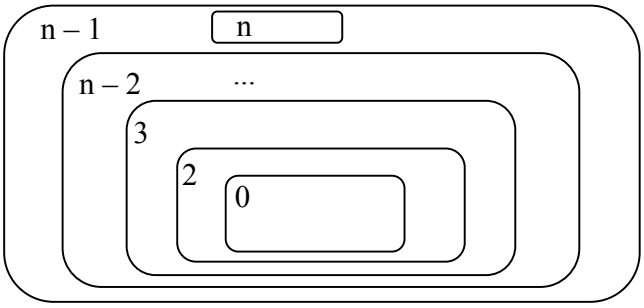
$$\Pi_4 = (V, \mu, A_0, A_2,..., A_n, (R_0, M_0),(M_2), ..., (M_{n-1}), \emptyset, i_0), \text{ kde}$$

$$V = \{ \langle i, r \rangle, [i, r] / 1 \le i \le n, 0 \le r \le n-1 \}$$
$$\cup \{ \langle i, r; j_1, ..., j_s \rangle \mid 1 \le i \le n, 0 \le r \le n-1, \{ j_1, ..., j_s \} \subseteq \{ j \in U \mid (i, j) \in E \} \}.$$

Objekty tvoří dva typy dvojic označení vrcholu i a označení pořadí r hrany vycházející z vrcholu a trojice určující kromě údajů i a r ve třetí komponentě také vrcholy, které jsou s vrcholem i spojeny hranou.

$$\mu = [_{n-1} [_n]_n [_{n-2} ... [_2 [_0]_0]_2 ...]_{n-2}]_{n-1},$$
$$A_0 = \{ (\langle 1,0 \rangle 1) \}, \quad A_i = \emptyset \text{ pro } i \ne 0$$
$$R_0 = \{ (\langle i, r \rangle \to [i, r] \langle j_1, r + 1 \rangle // \langle i, r; j_2, ..., j_s \rangle; zde, zde),$$
$$(\langle i, r; j_h, ..., j_s \rangle \to [i, r] \langle j_h, r + 1 \rangle // \langle i, r; j_{h+1}, ..., j_s \rangle; zde, zde),$$
$$(\langle i, r; j_{s-1}, ..., j_s \rangle \to [i, r] \langle j_{s-1}, r + 1 \rangle // \langle i, r; j_s, r + 1 \rangle; zde, zde) \mid$$
$$\text{pro } 1 \le i \le n - 1, 0 \le r \le n - 2, \text{ kde } j_1, ..., j_s \text{ jsou všechny } (i, j_l) \in E, 1 \le l \le s \},$$
$$M_0 = \{ (\langle n, n - 1 \rangle \to [n, n - 1]; ven) \},$$
$$M_i = \{ ([i, j] \to [i, j]; ven) \mid 1 \le j \le n - 1 \}, \text{ pro } 2 \le i \le n - 2,$$

$M_{n-1} = \{([n, n-1] \rightarrow [n, n-1]; do_n)\}$,



Obr. 6. Membránová struktura Π_4

Výpočet začíná z jednoho objektu *⟨1,0⟩*, který je umístěný v membráně označené *0*. Opakovaným použitím pravidla replikace se vždy prodlouží řetězec, který vlastně představuje cestu v grafu z vrcholu *1*. Žádny řetězec není dokončen dokud nedospěje k uzlu *n* nebo neprojde *n - 1* kroky. Tímto způsobem vytvoříme všechny cesty, které začínají v uzlu *1* a mají délku nejvíce *n - 1*.

Jen řetězce, které končí *⟨n, n - 1⟩*, mohou projít membránou *0*.

Každou další membránou *i*, $2 \leq i \leq n-2$, řetězec může (nezměněný) projít jen tehdy, obsahuje-li podřetězec *[i, t]*, pro nějaké $1 \leq t \leq n-1$. To znamená, že jsme tímto uzlem prošli v *t*-tém kroku. Pokud se řetězec dostane do oblasti ohraničené obalem, je poslán do výstupní membrány jen tehdy pokud obsahuje symbol *[n, n-1]*, což znamená, že cesta končí v uzlu *n*.

Máme tu odpověď na otázku, jestli má nebo nemá "Problém hamiltonovské cesty" řešení a tato cesta je uložena ve výstupní membráně *n*.

Věta [26]: $NWP_6 = RE$.

6 P automaty

P automat je membránová struktura tvořená hierarchicky vnořenými membránami. Membrána ohraničuje oblast obsahující multimnožinu objektů, pravidla a může obsahovat i další membrány. Vzájemná komunikace mezi membránami P automatu je realizována přechodem objektů mezi jednotlivými membránami. Nové objekty z prostředí vstupují do systému jen přes hlavní membránu. Objekty se během výpočtu nemění, pouze procházejí membránami. U původního modelu P automatu se objekty přesouvají jen jedním směrem, dovnitř oblasti. Přesouvání jednotlivých objektů je určeno pravidly, která jsou pevně svázána s jednotlivými oblastmi.

Obsahy jednotlivých oblastí na začátku výpočtu určují počáteční stav P automatu. Výpočet je ukončen, pokud se P automat dostane do koncového stavu, který je určen stejně jako počáteční stav obsahem oblastí. Výsledkem výpočtu je posloupnost multimnožin objektů, které byly přijaty systémem v průběhu úspěšného výpočtu, to znamená během takového výpočtu, kdy automat začal ve vstupním stavu a skončil v koncovém stavu. Použijeme-li dvě abecedy v definici P automatu a to abecedu pro vstupní symboly a abecedu

pomocných symbolů, bude vstup automatu tvořen pouze ze symbolů vstupní abecedy. V daném případě hovoříme o rozšířeném P automatu.

Je-li v každém kroku výpočtu v každé membráně použito pouze jedno pravidlo nedeterministicky vybrané z použitelných pravidel, hovoříme o sekvenčním P automatu [13]. Pokud jsou použita všechna použitelná pravidla v daném kroku, jedná se o P automat s maximálním paralelismem. Jsou-li k jednotlivým pravidlům přiřazeny priority, nazýváme daný automat sekvenční P automat s prioritami [5, 6, 8] nebo P automat s maximálním paralelismem s prioritami.

Uvedeme formální definici P automatu s prioritami.

Definice 5: P automat s prioritami a s n-membránami je struktura

$$\Gamma = (V, \mu, (w_1 , P_1, R_1, F_1), \dots , (w_n , P_n, R_n, F_n)), n \geq 1, \text{ kde}$$

- V je abeceda objektů.
- μ je membránová struktura složená z n membrán.
- $w_i \in V^0$, $1 \leq i \leq n$ je počáteční obsah oblasti i.
- P_i, $1 \leq i \leq n$, je konečná množina komunikačních pravidel oblasti i.
- R_i relace částečného uspořádání nad množinami P_i, $1 \leq i \leq n$, které specifikují priority mezi pravidly z množiny P_i.
- $F_i \subseteq V^0$, $1 \leq i \leq n$ je konečná množina koncových stavů (multimnožin) nebo $F_i = \emptyset$, pak $F_j \neq \emptyset$ pro alespoň jedno $j \neq i$.

P automat je P systém se speciálním typem pravidel. Komunikační pravidla mají tvar $(y, in)_x$ kde $x, y \in V^+ \cup \{\varepsilon\}$. Pro $x, y \neq \varepsilon$ řetězce x a y reprezentují multimnožinu $x, y \in V^0$. Pravidlo $(y, in)_x$ znamená: Pro $x, y \neq \varepsilon$, jestliže x je obsaženo v oblasti i a y je obsaženo v oblasti rodičovské membrány, pak multimnožina y přejde přes membránu do oblasti i. Jestliže $x = \varepsilon$, pak pro aplikaci pravidla oblast i musí být prázdná a jestliže $y = \varepsilon$, pak pravidlo nežádá žádnou multimnožinu z oblasti rodičovské membrány. Jedná se tedy o typ symport pravidel, ke kterým přidáme promoter neboli podmínku aplikovatelnosti. Pro jeden řetězec x může být s jednou membránou asociovaných více pravidel $(y, in)_x$ a $(z, in)_x$, takových že $y \neq z$.

V každém kroku výpočtu každá oblast žádá svou rodičovskou oblast (obal žádá okolí) o multimnožinu objektů na základě jednoho pravidla, které je nedeterministicky vybráno ze všech aplikovatelných pravidel. Jestliže jsou všechny požadavky splněny, pak se systém nachází v nové konfiguraci, která je určena obsahem oblastí. Jestliže nějaká pohledávka nemůže být vyřízena nebo není pravidlo, které by mohlo být aplikováno, systém výpočet neúspěšně ukončí. P automat se zastaví v konfiguraci, která není koncová.

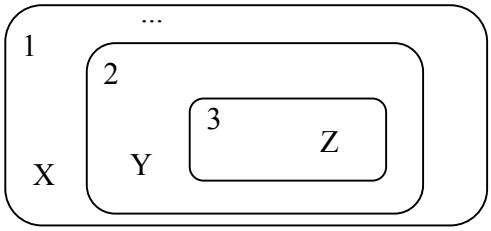
Příklad 4 [13]: Mějme P automat definovaný následovně:

$\Gamma = (\{X, Y, Z, a, b, c\}, [_1 [_2 [_3]_3]_2]_1, (X, P_1, \emptyset, \{\varepsilon\}), (Y , P_2, \emptyset, \{X, Y\}, (Z, P_3, \emptyset, \emptyset)),$

$P_1 = \{(a, in)_X, (a, in)_a, (b, in)_a, (b, in)_b, (c, in)_b, (c, in)_c, (\varepsilon, in)_c, (\varepsilon, in)_\varepsilon,$

$P_2 = \{(X, in)_Y, (a, in)_X, (b, in)_X, (c, in)_X, (c, in)_X, (\varepsilon, in)_c,$

$P_1 = \{(\varepsilon, in)_Z, (abc, in)_Z.$



Obr. 7. Struktura P automatu akceptujícího jazyk *L*

Nyní popíšeme činnost tohoto P automatu. Automat začíná výpočet v počáteční konfiguraci: (X, Y, Z).

S první oblastí je asociované jediné aplikovatelné pravidlo $(a, in)_x$, ve druhé oblasti je to pravidlo $(X, in)_Y$ a ve třetí oblasti je aplikovatelné pouze pravidlo $(\varepsilon, in)_Z$. Provedením těchto pravidel přejde P automat z počáteční do nové konfigurace. Po prvním kroku výpočtu je P automat v konfiguraci:

(a, XY, Z).

Nyní jsou v první oblasti aplikovatelná dvě pravidla $(a, in)_a$ a $(b, in)_a$. Systém nedeterministicky vybere jedno z nich, nechť je to první z nich $(a, in)_a$. Ve druhé oblasti je aplikovatelné pravidlo $(a, in)_X$ a ve třetí oblasti bude aplikováno pravidlo $(\varepsilon, in)_Z$. Konfigurace po druhém kroku výpočtu je:

(a, XYa, Z).

Výpočet může pokračovat použitím stejné trojice pravidel. Nechť výpočet pokračuje stejným způsobem po *i-1* kroků. Konfigurace P automatu po těchto krocích je:

(a, XYa^i, Z).

V dalším kroku je aplikovatelná následující trojice pravidel:

$(b, in)_a$, $(a, in)_x$, $(\varepsilon, in)_z$.

Provedení těchto pravidel převede P automat do nové konfigurace. V tomto kroku provedení pravidla $(b, in)_a$ ukončilo přijímání symbolů *a* do systému. V dalších krocích budou do systému vstupovat pouze symboly *b* a *c*. Nová konfigurace je:

(b, XYa^{i+1}, Z).

V první oblasti jsou aplikovatelná opět dvě pravidla, pravidlo $(b, in)_b$ a pravidlo $(c, in)_b$. První z nich zaručí přijetí dalších symbolů *b* do systému zatímco druhé pravidlo zastaví příjem symbolů *b* vstupem prvního symbolu *c* do systému. Předpokládejme tedy v následujících *j* krocích aplikaci trojice pravidel:

$(b, in)_a$, $(b, in)_x$, $(\varepsilon, in)_z$.

Po provedení těchto kroků výpočtu bude P automat v konfiguraci:

$(b, XYa^{i+1}b^j, Z)$.

V následujícím kroku použijeme trojici pravidel:

$(c, in)_b$, $(b, in)_x$, $(\varepsilon, in)_z$.

Systém má nyní konfiguraci:

$(c, XYa^{i+1}b^{j+1}, Z)$.

V této fázi výpočtu bude systém přijímat symboly *c*. Nechť je aplikována trojice pravidel: $(c, in)_b$, $(c, in)_x$, $(\varepsilon, in)_z$.

Aplikace pravidel převede systém do konfigurace:

$(c, XYa^{i+1}b^{j+1}c, Z)$.

Protože ve třetí oblasti jsou aplikovatelná již dvě pravidla, může výpočet probíhat více způsoby. V automatu teď probíhají dvě odlišné činnosti: (1) P automat přijímá určitý počet symbolů c, (2) do třetí oblasti jsou přesouvány trojice symbolů abc. V této oblasti je pak v každém momentu výpočtu stejný počet symbol a, b a c. Pravidla $(\varepsilon, in)_c$ a $(\varepsilon, in)_\varepsilon$ pomáhají vyprázdnit první oblast a vykonávat v první oblasti „prázdné“ kroky. Ke stejnému účelu slouží i pravidlo $(\varepsilon, in)_c$ asociované s druhou oblastí. Výpočet úspěšně skončí pouze tehdy, přijal-li systém stejný počet všech symbolů. Množina koncových stavů je definována tak, že (1) $F_3 = \emptyset$, při dosažení koncové konfigurace tedy nezáleží na obsahu třetí membrány; (2) $F_1 = \{\varepsilon\}$, což znamená, že pro dosažení koncové konfigurace musí být oblast první membrány prázdná; (3) $F_2 = \{X, Y\}$ a tedy druhá membrána může obsahovat při dosažení koncové konfigurace pouze symboly X a Y a žádné jiné symboly, které vstoupily do systému během výpočtu. Vzhledem k tomu, že jediné pravidlo, které může způsobit vyprázdnění druhé membrány, je pravidlo, které přesunuje vždy trojici symbolů abc do oblasti třetí membrány, pak daný P automat akceptuje jazyk $L = \{a^i, b^i, c^i \mid i \geq 1\}$.

Bylo dokázáno, že P automat s prioritami a dvěma membránami má výpočetní sílu RE.

Věta [6]: Každý rekurzivně vyčíslitelný jazyk $L \subseteq \Sigma^*$ můžeme vyjádřit jako jisté zobrazení jazyka P automatu s prioritami se dvěma membránami.

Stejně jako u klasického automatu jsou u P automatu s prioritami definované počáteční a koncový stav, konfigurace, přechodová funkce, výpočet a přijímaný jazyk. Proto také stejně jako u tradičního automatu byl navržen algoritmus minimalizace P automatu obecně n membránového na dvoumembránový [8]. Byl studován problém vzájemného převodu sekvenčního P automatu a P automatu s maximálním paralelismem. Přímou transformaci sekvenčního P automatu na automat s maximálním paralelismem a naopak můžeme najít v [2]. Pro danou přímou transformaci byl definován rozšířený P automat s prioritami, kde je původní definice rozšířena o množinu terminálních symbolů. V práci [4] jsou porovnány jednotlivé typy P automatů.

7 P kolonie

Nejmladší P strukturou jsou P kolonie, kolonie jednomembránových struktur, které byly zavedeny v práci [22] v roce 2004. Struktury jsou umístěné ve společném prostředí, ve kterém mají k dispozici speciální elementární objekty e. V případě, že si membrána symbol e z prostředí vyžádá, je tento symbol vždy k dispozici. Prostředí P kolonií slouží jako komunikační kanál i „skladiště“ objektů. Počet objektů uvnitř každé membrány je stejný a vzhledem na deklarovanou jednoduchost membránové struktury i malý a v průběhu výpočtu se nemění. Činnost každé jednomembránové struktury je řízena vlastními pravidly strukturovanými do programů. Počet pravidel programu se shoduje s počtem objektů uvnitř membrány a zabezpečuje akci (změnu nebo přesun) všech objektů uvnitř membrány v každém diskrétním okamžiku na základě zvoleného programu. Výsledek výpočtu je dán počtem speciálních symbolů v prostředí po zastavení. Zavedeme P kolonie formálně.

Definice 6: P kolonie Π s kapacitou c je struktura

$$\Pi = (A, e, f, B_1, \ldots, B_n),$$ kde

- A je abeceda kolonie (prvky A jsou objekty).

- $e \in A$ je základní objekt kolonie.
- $f \in A$ je koncový objekt kolonie.
- $w_E \in A^*$ reprezentuje počáteční multimnožinu v prostředí.
  B_i, $1 \leq i \leq n$ jsou elementární membránové struktury $B_i = (O_i, P_i)$, kde
  - O_i je multimnožina c objektů a
  - $P_i = \{ p_{i,1}, ..., p_{i,k_i} \}$ je konečná množina programů;

  každý program $p_{i,j}$ sestává z c pravidel: vývojových $a \rightarrow b$, kde $a,b \in O_i$, komunikujících $c \leftrightarrow d$, pro $c,d \in O_i$, a/nebo kontrolních $c \leftrightarrow d / c' \leftrightarrow d'$, kde $c, d, c', d' \in O_i$.

Činnost jednomembránové struktury: Vývojová pravidla $a \rightarrow b$ přepisují objekty uvnitř membrány, pravidla komunikace $c \leftrightarrow d$ zabezpečují výměnu objektů mezi membránou a prostředím. Dvě pravidla lze spojit do kontrolního pravidla např. $c \leftrightarrow d / c' \leftrightarrow d'$. Na přepisu se podílí jenom jedno z pravidel v kontrolním pravidle. Přednost má první z nich, druhé přepisuje jenom když nelze použít první pravidlo. V jednom kroku přepisuje objekty v membráně některý z programů, nedeterministicky vybraný z těch programů, které umožňují zpracovat aktuální objekty membrány. Počet programů agenta není omezen.

Analogicky jako obecně v P systémech, i u stávajícího modelu nás zajímá výpočetní síla při sekvenční i paralelní činnosti membrán, dále možnosti P kolonií ve vztahu k počtu objektů v membránách i v závislosti na skladbě pravidel v programech, i na specifikaci začátku výpočtu. Výsledkem výpočtu je počet kopií speciálního symbolu f v prostředí po ukončení práce kolonie. P kolonii přiřadíme tři konstanty: c - počet objektů v membráně, n - počet membrán a h - maximální počet programů v membráně.
Budeme se zabývat třídami kolonií $NPCOL_m\alpha(c,n,h)$, kde
- $m \in \{par, seq\}$ odlišuje paralelně nebo sekvenčně pracující (membrány) P kolonie
- $\alpha \in \{\varepsilon, K, R, KR\}$ představuje následující omezení na tvar programů: ε programy bez kontrolních pravidel, K programy s kontrolními pravidly, R programy ve tvaru $< a \rightarrow b, c \leftrightarrow d >$ a $< a \rightarrow b, c \leftrightarrow d / c' \leftrightarrow d' >$.

Případ, kdy hodnoty parametrů n nebo h nebudou omezeny bude vyznačen symbolem *.
Následující věty sumarizují výsledky dosažené v této oblasti pracích [9, 14, 17, 20]:

Věta: $NPCOL\ KR(2,1,*) = NRE$.
$NPCOL\ (2,1,*) = NMAT$ [3]
$NPCOL_{par}\ (1,*,7) = NRE$
$NPCOL_{par}\ (1,4,*) = NRE$
$NPCOL_{seq}KR(2,*,5) = NRE$
$NPCOL_{par}KR(2,*,5) = NRE$
$NPCOL_{par}R(2,*,5) = NRE$
$PCOL_{par}\ K(2,*,4) = NRE$
$NPCOL_{par}\ K(3,*,3) = NRE$

8 Závěr

Cílem práce bylo podat základní informaci a informaci o trendech vývoje nové a rychle se vyvíjející oblasti výzkumu o membránových systémech. Čtenáře, kterého membránové počítání zaujalo odkazujeme na bohatou literaturu v této oblasti.

[3] *NMAT* množiny čísel generované maticovými gramatikami

Literatura

[1] Baranda, A. V., Castellanos, J., Arroyo F., Luengo, C.: Bio-Language for Compu-ting with Membranes. In: J. Kelemen, P. Sosik (eds.): *Advances in Artificial Life.* Berlin: Springer-Verlag 2001, pp. 176-185.

[2] Cienciala, L.: P automata with priorities working in sequential and maximally parallel mod. In: G. Paun, C. Zandron, G. Mauri (eds.): *Pre-Proceedings of the Worskhop on Membrane Computing* Milano, 2004, pp. 186-195.

[3] Cienciala, L.: Rozšířený P automat. In: J. Kelemen, V. Kvasnička (eds.): *Kognice a umělý život IV*. Opava: Slezská univerzita 2004, pp. 95 - 105.

[4] Cienciala, L.: P automaty. In: J. Kelemen, V. Kvasnička, J. Pospíchal, (eds.) *Kognice a umělý život V*. Opava: Slezská univerzita 2005, pp. 103 - 115.

[5] Cienciala, L., Ciencialová, L.: Membranové automaty. In: J. Kelemen (ed.): *Kognice a umělý život III.* Opava: Slezská univerzita 2003, pp. 33 - 44.

[6] Cienciala, L., Ciencialová, L.: P automata with priorities. In: A. Alhazov, C. Martin-Vide, G. Paun (eds.): *Pre-Proceedings of the workshop on membrane computing.* Tarragona 2003, pp. 161 - 168.

[7] Cienciala, L., Ciencialová, L.: Sekvenční a paralelní P automaty. In: J. Kelemen, V. Kvasnička (eds.): *Kognice a umělý život IV* (Opava: Slezská univerzita 2004, pp. 107 - 116.

[8] Cienciala, L., Ciencialová, L.: Membrane automata with priorities. *Journal of Computer Science and Technology* **1** (2004) 89-97, Special Issue on Bioinfomatics (P. Bonizzoni, G. D. Vedova, T. Jiang, eds.).

[9] Ciencialová, L., Cienciala, L.: Variations on the Theme: P colonies. In: D. Kolář, A. Meduna (eds.): *Proceedings of 1st International Workshop WFM '06.* Ostrava, 2006, str. 27-134. (also available on-line as a CEUR Workshop Proceedings - http://CEUR-WS.org).

[10] Ciencialová, L., Cienciala, L.: P kolonie. In: J. Kelemen, V. Kvasnička (eds.): *Kognice a umělý život VI.* Opava: Slezská univerzita 2006, pp. 119-125 .

[11] Csuhaj-Varjú, E., Dassow, J., Kelemen, J., Păun, G.: *Grammar Systems – A Grammatical Approach to Distribution and Cooperation*. Yverdon: Gordon and Breach 1994.

[12] Csuhaj-Varjú, E., Kelemen, J., Kelemenová, A., Păun, G., Vaszil, G.: *Cells in environment: P Colonies*, *Journal of Multi-Valued Logic* (to appear).

[13] Csuhaj-Varjú, E., Vaszil, G.: P Automata. In: G. Păun, G. Rozenberg, A. Salomaa, C. Zandron (eds.): *Membrane Computing*. LNCS 2597. Berlin: Springer 2003, pp. 219-233.

[14] Csuhaj-Varjú, E., Margenstern, M., Vaszil, G.: P Colonies with a Bounded Number of Cells and Programs. In: Hoogeboom, H. J., Păun, Gh., Rozenberg, G., Salomaa, A. (eds.): *Membrane Computing, 7th International Workshop.* LNCS 4361, Berlin: Springer 2006, pp. 352-366.

[15] Dassow, J., Păun, Gh.: On the Power of Membrane Computing. *J. of Universal Computer Sci.* **2** (1999) 33-49.

[16] Freund, R.: Special variants of P systems inducing an infinite hierarchy with respect to the number of membranes. *Bulletin of the EATCS* **75** (2001) 209-219.

[17] Freund, R., Oswald, M.: P colonies working in the maximally parallel and in the sequential mode. In: G. Ciobanu, Gh. Păun (eds): *Pre-Proceedings of the 1st International Workshop on Theory and Application of P Systems.* Timisoara, Romania 2005, pp. 49-56.
[18] Kelemen, J., Kelemenová, A.: Život ve světě symbolů: úvod do jazykověteoretického studia živosti. In: J. Kelemen, V. Kvasnička, J. Pospíchal (eds.): *Kognice a umělý život* (Opava: Slezská univerzita 2001, pp. 101-118.
[19] Kelemen, J., Kelemenová, A: On P colonies, a simple bio-chemically inspired model of computation. In: I. J. Rudas (ed.): *Proc. 6th International Symposium on Computational Intelligence.* Budapest: Budapest Tech. 2005, pp. 40-56.
[20] Kelemen, J., Kelemenová, A., Păun, Gh.: *On the power of a biochemically inspired simple computing model*: *P colonies*.
[21] Downoloadable version at http://psystems.disco.unimib.it.
[22] Kelemen, J., Kelemenová, A., Păun, G.: The power of cooperation in a biochemically inspired computing model: Preview of P colonies: In: M. Bedau at all, (eds.): *Workshop and Tutorial Proceedings, 9th International Conference on the Simulation and Synthesis of Living Systems*, *ALIFE IX.* Cambridge, MA: MIT Press 2004, pp. 82-86.
[23] Krishna, S. N., Rama, A.: A note on parallel rewriting in P systems. *Bulletin of the EATCS* **73** (2001) 147-151.
[24] Marcus, B.: Bridging P Systems and Genomics: A Premilitary Approach. In [23].
[25] Markoš, A.: *Tajemství hladiny.* Praha: Vesmír 2000.
[26] Martin-Vide, C., Mitrana, V.: P systems with valuations. In: I. Antoniou, C. S. Calude, M. J. Dinneen (eds.): *Unconventional Models of Computation* (London: Springer-Verlag 2000, pp. 154-166.
[27] Martin-Vide, C., Păun, Gh.: Computing with membranes. One more collapsing hierarchy, *Bulletin of the EATCS* **72** (2000) 183-187.
[28] Martin-Vide, C., Păun, Gh.: Language generating by means of membrane systems. *Bulletin of the EATCS* **75** (2001) 199-208.
[29] Păun, Gh.: Computing with membranes: An Introduction. *Bulletin of the EATCS* **67** (1999) 139-152.
[30] Păun, Gh.: Computing with membranes. *Journal of Computer System Sciences* **61**,1 (2000) 108-143.
[31] Păun, Gh.: From Cells to Computers: Computing with membranes (P Systems). R. Freund, A. Kelemenová (eds.): *Proceedings of the International Workshop Grammar Systems 2000.* Opava: Silesian University 2000, pp. 9-40.
[32] Păun, Gh.: *Membrane computing. An introduction.* Berlin-Heidelberg: Springer Verlag 2002.
[33] Păun, A., Păun, M.: On the membrane computing based on splicing. In: C. Martin-Vide, V. Mitrana (eds.): *Where Mathematics, Computer Science, Linguistics and Biology meet*. Dordrecht: Kluwer 2001, pp. 409-422.
[34] Păun, Gh., Rozenberg, G., Salomaa, A.: *DNA Computing. New Computing Paradigms*. Heidelberg: Springer-Verlag 1998.
[35] Păun, Gh., Rozenberg, G., Salomaa, A.: Membrane computing with external output. *Fundamenta Informaticae* **41**, 3 (2000), 259-266.

[36] Rodriguez-Paton, A.: On the universality of P systems with membrane creation. *Bulletin of the EATCS* **74** (2001) 229-234.
[37] The P systems web page: http://psystems.disco.unimib.it.
[38] Thomas, L.: Život buňky. In: L. Thomas: *Myšlenky pozdě v noci.* Praha: Mladá fronta 1989.
[39] Zandron, C., Mauri, G., Ferretti, C.: Universality and normal forms on membrane systems. In: R. Freund, A. Kelemenová (eds.): *Proceedings of the International Workshop Grammar Systems 2000.* Opava: Silesian University 2000, pp. 61-74.

Vplyv mutácie na štruktúru organizmov

Marián Ivančo[1]

Abstrakt. Pri vysokej mutácii sa v populácii presadzujú tie organizmy, u ktorých zmena vo vnútornej štruktúre spôsobuje najmenší pokles schopností. Pri stredných veľkostiach mutácie môžu organizmy dočasne obetovať nejakú vlastnosť pre budúce získanie inej. Pri nízkej mutácii by pokles schopností viedol k vytlačeniu organizmu z populácie ešte pred získaním novej výhody. Mutácia teda zohráva aktívnu úlohu v smerovaní vývoja organizmov v prostredí.

1 Úvod

Rast zložitosti organizmov v prírode patrí medzi kontroverzné témy možno práve pre absenciu vhodnej miery zložitosti. Organizmus ako taký sa ťažko exaktne popisuje vzhľadom na problémy s definíciou života. Pravdepodobne sa zhodneme, že baktéria E. coli je jednoduchším predstaviteľom života ako napríklad kukurica. Do obtiaží sa dostaneme, ak chceme porovnať zložitosť kukurice a myši. Ako zistíme, ktorý organizmus je zložitejší a akým spôsobom táto zložitosť vznikla? Môžeme skúmať „zápis“ (genotyp) alebo „prejav“ (fenotyp) organizmu. Ak sa rozhodneme skúmať fenotypy, dostávame sa do problémov ako určiť podľa akých kritérií budeme porovnávať organizmy. Je dýchanie zložitejší prejav ako trávenie? Alebo ako musíme prejavy deliť na atomárne procesy?

Dostávame sa k problému určenia, ktoré činnosti sú už tie „pravé“ nedeliteľné a ako ich navzájom porovnať. Vzhľadom na rôznorodosť a nepredvídateľnosť prejavov života bude asi ťažké určiť takúto množinu. Do podobných problémov sa dostaneme, ak budeme porovnávať fyzickú zložitosť, alebo zloženie organizmu z funkčných častí. Treťou, asi najschodnejšou cestou, je porovnanie genómu alebo zápisu organizmu. Je asi prirodzené predpokladať, že organizmus s viacerými prejavmi alebo schopnosťami bude mať aj zložitejší zápis. Nesmieme ale zabúdať na nevyužité alebo neaktívne miesta v genóme organizmu. Tieto miesta síce plnia dôležitú úlohu v evolúcii, nepredstavujú ale z hľadiska zložitosti žiadny prínos. Extrémnym príkladom je náhodný reťazec, ktorý v prevažnej väčšine prípadov nepredstavuje živý organizmus, ale jeho vnútorná štruktúra je na pohľad „veľmi“ zložitá. Na opačnom konci spektra je reťazec zložený len z rovnakých znakov, napr. „aaaaa“, ktorý vieme jednoducho popísať alebo charakterizovať. O oboch takýchto reťazcoch môžeme predpokladať, že majú nulovú biologickú zložitosť, keďže nemajú prakticky žiadnu biologicky zaujímavú vlastnosť. Nedokážu sa rozmnožovať ani inak sa navonok zmysluplne prejavovať. Pozrime sa preto ako môžeme skúmať zložitosť organizmu a ako sa menia jeho vlastnosti v premenlivom prostredí. Pre naše pokusy budeme využívať simulačné prostredie Avida.

[1] Fraňa Kráľa 56, 903 01 Senec, E-mail: marian.ivanco@gmail.com

2 Prostredie Avida

Na pokusy budeme používať program Avida, ktorý vznikol v roku 1998, a za ktorým stáli Chris Adami, Charles Ofria a Richard Lenski [4]. Avida bola inšpirovaná programom Tierra a mala simulovať evolúciu asexuálnej populácie baktérií. Pri asexuálnom rozmnožovaní nedochádza ku kríženiu a k rekombinácii. Nové vlastnosti sa šíria pomalšie ako v sexuálnej populácii, čo je ešte zvýraznené spôsobom, akým sú potomkovia vkladaní do prostredia. Zároveň to predlžuje čas prežitia menej schopných organizmov, čo môže ovplyvniť evolúciu hlavne pri menších mutáciách. Ďalšou vlastnosťou prostredia je sekvenčné vykonávanie inštrukcií, ktoré ovplyvňuje rozloženie genetickej informácie v genóme ako aj spôsob vzniku nových vlastností.

Organizmy sú v Avide reprezentované sadou jednoduchých inštrukcií napísaných na páske a predstavujú akýsi jednoduchý program (podobne ako inštrukcie programu v pamäti počítača). Organizmy navzájom priamo neinteragujú, ale sa nepriamo vytláčajú z populácie, ktorej veľkosť je určená veľkosťou mriežky, teda prostredia. Pri rozmnožovaní sa jedinec rozdelí na dve časti, pričom novovzniknutá kópia vytlačí z okolia (mriežky 3×3) iného náhodného jedinca. Nové vlastnosti sa teda šíria v prostredí pomalšie (kvadraticky) ako keby sa potomok vložil kdekoľvek do prostredia (exponenciálne šírenie). Úlohou programu je skopírovať seba samého a rozdeliť sa.

Spočiatku sa všetky organizmy v populácii vykonávajú rovnako rýchlo. Podobne ako v prírode získavajú baktérie energiu z chemických reakcií, môže organizmus v Avide získať energiu (teda zrýchliť svoje vykonávanie) vykonaním jednoduchých bitových operácií. Úlohou organizmu je načítať jedno alebo viac čísel z prostredia a vrátiť nové číslo. Prostredie nekontroluje ako sa organizmus k výsledku dopracoval a organizmus nemusí označiť o akú operáciu ide. Ak povolíme napr. bitové operácie AND, OR alebo XOR a organizmus načíta čísla 1100 a 0101, prostredie len skontroluje, či je vrátené číslo zhodné s 0100 (= 1100 AND 0101), 1101 (= 1100 OR 0101) alebo 1001 (= 1100 XOR 0101). Ak áno, odmení organizmus zrýchleným vykonávaním jeho programu v závislosti od zložitosti vykonanej operácie.

Avida nám poskytuje služby ako napríklad možnosť ukladať jednotlivé generácie a pri vhodnom nastavení aj presne zopakovať pokus (aj s náhodnými zmenami). Nesmieme ale zabúdať aj na nevýhody. Organizmus musí byť napríklad schopný rozmnožovania, inak Avida nie je schopná zhodnotiť schopnosti organizmu a z tohto dôvodu nie je jednoduché oddeliť inštrukcie rozmnožovania od iných inštrukcií. Toto komplikuje neskoršiu automatickú analýzu štruktúry genetickej informácie.

3 Sledovanie zmien a odhad zložitosti

V našom pokuse budeme okrem iného sledovať tri vlastnosti genómu. Jeho modularitu, teda spôsob, akým sú rozložené relevantné inštrukcie v genóme a či sú organizované v zhlukoch. Ďalej recykláciu, teda či a ako dokáže organizmus po zmene prostredia znovu použiť dovtedy relevantné inštrukcie a štruktúry. A nakoniec preťažovanie, teda využívanie jednej inštrukcie na viacero úloh. Pri skúmaní recyklácie budeme musieť skúmať organizmus po prvej časti evolúcie ako aj po zmene prostredia. V druhej časti budeme skúmať zložitosť organizmu meranú informačnou zložitosťou genómu a to viacerými spôsobmi, ktoré nám pomôžu vytvoriť celkový obraz o genóme.

Modularita genómu je typický jav pri sexuálnom rozmnožovaní, pri ktorom dochádza ku kríženiu. Hlavnou hnacou silou pre vznik modularity je práve kríženie a s ním súvisiace riziko rozdelenia, a teda aj znefunkčnenia dlhších génov. Pri asexuálnom rozmnožovaní mechanizmus kríženia absentuje. *Preťažovanie* je tiež typická vlastnosť zložitejších organizmov. Katalyzátorom jej vzniku môže byť málo priestoru, krátky genóm v porovnaní s počtom a zložitosťou očakávaných vlastností. Organizmus bude v takom prípade nútený využívať časti svojho genómu rôznymi spôsobmi. Iným mechanizmom môže byť frekvencia chýb pri rozmnožovaní, teda mutácia. Ak je veľkosť mutácie malá, teda vznikne málo zmien v potomkovi, bude cena vzniku novej vlastnosti vyjadrená potrebným časom relatívne vysoká. Lacnejšou cestou bude v tomto prípade prispôsobenie si už existujúcej štruktúry v genóme aj na iné potreby. Veľkosť mutácie môže mať vplyv aj na *recykláciu*. V tomto prípade je cena potrebná na „zápis“ novej vlastnosti oveľa vyššia ako úprava už existujúcej štruktúry v genóme. Po skončení evolúcie v novom prostredí môže byť organizácia genómu, aj s prihliadnutím na preťažovanie, veľmi previazaná. Organizmus v tomto prípade nemôže prepísať alebo upraviť časť kódu, pretože by poškodil aj iné vlastnosti a bude nútený skôr nepriamo bez zmeny aktívnych častí pozmeniť celkové správanie sa. Ide tu vlastne o akési nabaľovanie kódu bez väčšej zmeny pôvodných častí.

Pre skúmanie týchto vlastností zostrojíme tzv. *minimálny genóm*, ktorý vznikne z pôvodného znefukčnením jednotlivých pozícií tak, aby sa výsledný organizmus správal rovnako ako pôvodný. Jednotlivé pozície nebudeme z genómu odstraňovať. ale nahradzovať prázdnymi inštrukciami. Dostaneme tak akúsi minimálnu kostru organizmu zodpovednú za správanie sa. Treba podotknúť, že táto kostra nie je jednoznačne určená.

Pravdepodobne najpoužívanejšia metóda na meranie *zložitosti* reťazca (genómu) je Kolmogorovova zložitosť. Je definovaná ako dĺžka najkratšieho programu schopného vygenerovať daný reťazec. Tento prístup však naráža na viaceré biologické nedostatky. Náhodné reťazce majú vysokú Kolmogorovovu zložitosť, keďže program generujúci daný náhodný reťazec musí vlastne v sebe obsahovať zápis danej sekvencie. Naopak reťazce popisujúce fraktály a podobné štruktúry majú nízku Kolmogorovovu zložitosť, pretože tieto štruktúry mávajú jednoduchý popis svojho generovania. Kolmogorovova zložitosť je z tohto dôvodu skôr vhodná ako popis pravidelnosti ako biologickej zložitosti. Niektoré nevýhody Kolmogorovovej miery zložitosti sa pokúsil odstrániť Bennett [3] vo svojej definícii logickej hĺbky. Na určenie zložitosti si všíma čas behu a nie dĺžku samotného programu. Komplexné, aj keď pravidelné reťazce (ako fraktály), majú potom väčšiu zložitosť. Problém s náhodnými reťazcami ale zostal.

Problém náhodných reťazcov je nutné vyriešiť aj preto, že podstatné časti genómu organizmu sa nepodieľajú na vlastnostiach organizmu, neovplyvňujú jeho šance na prežitie v stabilnom prostredí a sú vlastne neaktívne v živote organizmu. Ich dôležitosť sa prejavuje hlavne pri rozmnožovaní, kedy do hry vstupuje mutácia ako motor evolúcie. Malou zmenou v neaktívnej časti môže organizmus získať novú dôležitú vlastnosť. Táto časť organizmu teraz môže zvýšiť zložitosť organizmu, ale dovtedy nijako podstatne neovplyvňuje organizmus. Dostávame sa teraz k dôležitej vlastnosti miery biologickej zložitosti. Zložitosť organizmu je možné objektívne zhodnotiť len vzhľadom na prostredie, v ktorom sa nachádza [2]. Bez prítomnosti prostredia nie sme schopní posúdiť, či daná časť genómu je náhodný reťazec, alebo či aktívne napomáha prežitiu.

3.1 Shannonova teória informácie

Shannonova teória bola pôvodne vyvinutá pre potreby prenosu informácie cez zašumený kanál. Hovorí nám koľko bitov informácie potrebujeme na prenos určitého množstva informácie cez informačný kanál. Nech p_i je pravdepodobnosť výskytu genómu s_i z populácie X, potom teória definuje entropiu, alebo neurčitosť či neusporiadanosť, ako

$$H(X) = - \Sigma_{i=1} p_i \log p_i$$

Ak by v prostredí nebola žiadna selekcia, potom by entropia populácie dosiahla maximum H_{max}. Pre organizmy s rovnakou dĺžkou L potom platí:

$$H_{max}(X) = \log 2^L = L$$

Evolúciu si teraz môžeme predstaviť ako prenos informácie o prostredí v genóme cez (mutáciou) zašumený kanál. Naša neistota o genóme v ďalšej generácii je potom:

$$H(X|Y) = - \Sigma_{x=1} \Sigma_{y=1} p(y)\, p(x|y) \log p(x|y)$$

Vzorec nám hovorí o neurčitosti výstupu za predpokladu, že poznáme vstup. Teraz si môžeme informáciu, ktorá sa dostala do ďalšej generácie, definovať ako

$$I(X:Y) = H(X) - H(X|Y)$$

V roku 2000 Adami a Cerf [6] navrhli tzv. fyzickú zložitosť, ktorá počíta zložitosť organizmu vzhľadom na prostredie pomocou populácie navonok rovnakých organizmov s rovnakou dĺžkou genómov L. Ide o modifikáciu Shannonovej teórie. Množstvo informácie, ktoré má populácia o prostredí (fyzická zložitosť), je dané naledovne

$$I(X:E) = H_{max} - H(X|E) = L + \Sigma_{i=1} p_i \log_D p_i$$

Nejde v tomto prípade o zložitosť priamo v Shannonovom zmysle, E tu nie je množina genómov, ale prostredie. Vypočítať H(X|E) sa však ukázalo ako komplikované, a preto Adami navrhol jej aproximáciu pomocou entropie jednotlivých pozícií v genóme. Nech

$$H_i = - \Sigma_{j=1} p_j(i) \log_D p_j(i)$$

je entropia na pozícii i. Hodnoty $p_j(i)$ jednoducho odhadneme zo zastúpenia v populácii. Treba ešte spomenúť, že logaritmus je pri základe D, čo je veľkosť abecedy. Potom

$$I(i) = H_{max} - H_i = n + \Sigma_{j=1} p_j(i) \log_D p_j(i)$$

$$I = -\Sigma_i I(i)$$

ak veľkosť abecedy je 2^n a celková fyzická zložitosť je teda daná sumou $I(i)$. Táto metóda na určenie zložitosti má aj svoje nevýhody. Problémom je takzvaný efekt únosu (hitchhiking effect) [2], ktorý spôsobuje nadhodnotenie zložitosti populácie po objavení sa novej výhodnej vlastnosti ešte pred dosiahnutím ekvilibria.

Huang, Ofria a Torng [1] navrhli spôsob, ako vylepšiť predošlú metódu. Metóda je založená na simulovaní všetkých jednobodových mutácií. Predstavme si populáciu, v ktorej sú povolené len genómy so vzdialenosťou jedna od študovaného genómu. Našou úlohou je odhadnúť $p_j(i)$, kde i je pozícia v genóme a j písmeno abecedy. Označme $\omega_k(i)$ ako fitnes genómu, kde na pozícii i bol zmenený znak na k (označme ho $k(i)$). Nech $p_k(i)$ je pravdepodobnosť jeho výskytu v populácii. Potom priemerná fitnes je

$$\omega(i) = \Sigma_{k=1} p_k(i)\, \omega_k(i)$$

a ak predpokladáme, že populácia je v ekvilibriu, potom musí zároveň platiť aj

$$p_j(i) = (p_j(i)\, \omega_j(i) / \omega(i))\, (1 - \mu) + \Sigma_{k=1} (p_k(i)\, \omega_k(i) / \omega(i))\, \mu / (1 / D)$$

kde μ je veľkosť mutácie a D je veľkosť abecedy. Prvý sčítanec vyjadruje prežitie genómu bez zmeny mutácie, druhý sčítanec zase vyjadruje podiel všetkých organizmov zmutovaných na genóm $j(i)$. Zjednodušením dostávame

$$p_j(i) = (p_j(i)\, \omega_j(i) / \omega(i))\, (1 - \mu) + \mu / (1 / D)$$

Ostáva nám experimentálne zistiť hodnoty $\omega_j(i)$. Pomocou týchto dát vytvoríme sústavu rovníc, ktorú vyriešime a získame pravdepodobnosti. Popísaná metóda má samozrejme tiež svoje limity. Prehliada možné vzájomné ovplyvňovanie sa jednotlivých pozícií v genóme. Na druhej strane metóda odstraňuje efekt únosu, ako aj nutnosť vytvárať veľkú populáciu.

V tomto článku použijeme na odhad zložitosti minimálny genóm zostrojený pre skúmanie predošlých vlastností. Tento odhad bude slúžiť ako minimálny odhad zložitosti genómu, keďže ide o minimálnu sadu inštrukcií s rovnakými fenotypickými vlastnosťami. Jednotlivé verzie minimálneho genómu, ktoré dostaneme postupným odstraňovaním inštrukcií, majú veľmi podobnú fitnes vzhľadom na rovnaké vlastnosti výsledného organizmu. Drobné odchýlky môže spôsobovať rýchlosť, akou organizmus stihne získať výhody. Výsledná fitnes nám aj tak nevraví nič o ekvilibriu populácie, keďže všetky varianty boli z jedného pôvodného genómu a neexistujú popri sebe. Na druhej strane si výsledné minimálne genómy môžeme predstaviť ako populáciu možných potomkov zachovávajúcich informáciu o zakódovaní požadovaného pôvodného správania sa. Úlohou takýchto potomkov je len zachovať schopnosť vykonať dané operácie v prostredí. Pýtame sa aký je informačný obsah tejto schopnosti, prípadne informácie o prostredí.

Na základe pravdepodobnosti jednotlivých variantov minimálneho genómu môžeme odvodiť celkovú entropiu genómu a z nej aj informačný obsah. Povedzme, že pravdepodobnosť získania varianty j minimálneho genómu je q_j. Potom pre entropiu na pozícii k platí

$$H_k = - \Sigma_{j=1}\, p_j(k) \log p_j(k)$$

$$p_j(k) = \Sigma_l\, q_l\, p_j^l(k)$$

kde $p_j^l(k)$ je pravdepodobnosť, že na mieste k bude vo variante l symbol j. Pri veľkosti abecedy 26 je potom výsledná zložitosť

$$I_{26}(k) = H_{max} - H_k = 1 + \Sigma_{j=1}\, p_j(k) \log_{26} p_j(k)$$

$$I = \Sigma_k\, I(k)$$

kde $I_{26}(k)$ je zložitosť určená pozične, alebo inak povedané v sústave, v akej je zapísaný genóm. Tento náš odhad $I(k)$ je mierne odlišný od hodnôt v predošlých dvoch metódach, pretože zanedbávame možnosť vplyvu niektorých inštrukcií na beh genómu na danej pozícii. Niektoré inštrukcie na danom mieste môžu samozrejme spôsobiť zaniknutie niektorej z vlastností a v najhoršom prípade aj neschopnosť organizmu rozmnožovať sa, a teda nulovú fitnes.

4 Pokus

Pokus rozdelíme na dve časti. V jednej budeme pozorovať povrchovú štruktúru genómu ako modularita, preťažovanie a recyklácia a v druhej časti budeme skúmať informačnú zložitosť. V prvej časti budeme postupovať v dvoch krokoch, v prvom kroku predtrénujeme populáciu aby získala určitú štruktúru a v druhom kroku po zmene prostredia sa bližšie pozrieme ako sa populácia vysporiadala s touto zmenou. V druhom prípade nám bude stačiť len jednokroková adaptácia na prostredie. Pokusné populácie si rozdelíme na tri skupiny. Prvú skupinu M_4 budeme trénovať na štyri operácie, menovite Not, Nand, OrNot a Or. Druhá skupina M_3 bude obsahovať tri operácie, menovite Nor, Xor a Equals. Tretia skupina M_7 bude trénovaná na množine všetkých siedmich operácií. Vo všetkých prípadoch budeme sledovať vývoj pri týchto troch úrovniach mutácie 0,025, 0,0025 a 0,00025.

4.1 Sledovanie povrchovej štruktúry genómu

S narastajúcim počtom inštrukcií a zložitosti narastá aj počet relevantných inštrukcií. Skúsme pokryť relevantné inštrukcie intervalmi. Ak použijeme viac intervalov, tieto pokryjú menšiu časť genómu. Snažíme sa čo možno najkratšími intervalmi obsiahnuť všetky relevantné inštrukcie. Ak sa nepokrytá časť výrazne nezvýši vieme odhadnúť počet zhlukov inštrukcií. V tabuľke 1 vidíme, že pre mutáciu 0,025 a M_4 je genóm tvorený pravdepodobne 4 zhlukmi relevantných inštrukcií. Prítomnosť zhlukov by bola evidentnejšia, ak by bol pomer relevantných inštrukcií a dĺžky genómu menší.

Preťažovanie inštrukcií vypovedá o pravdepodobnosti ovplyvnenia viacerých vlastností organizmu naraz. Veľkosť preťažovania napovedá ako efektívne je organizmus schopný (nútený) využívať priestor v genóme. Na rozdiel od bežných programov, ktoré vytvárajú ľudia, obsahuje genóm aj „nevyužité“ miesta. Čím viac takéhoto voľného priestoru má genóm k dispozícii, tým väčšia je šanca na získanie novej vlastnosti. Stačí si uvedomiť, že v dostatočne dlhom reťazci náhodných znakov sa s vysokou pravdepodobnosťou nachádza ľubovoľný pevne daný reťazec.

Tabuľka 1. Modulárne vlastnosti genómu pre rôzne prostredia a mutácie.

mutácia	prostredie	počet	priemer	pokrytie inštrukcií		
				3	4	5
0,02500	M_4	23	30	35 %	30 %	28 %
	M_3	34	31,88	56 %	50 %	45 %
	M_7	44	35,79	77 %	71 %	66 %
0,00250	M_4	25	26,32	47 %	33 %	28 %
	M_3	33	32,12	54 %	49 %	44 %
	M_7	48	41,9	85 %	81 %	77 %
0,00025	M_4	36	48,69	78 %	69 %	61 %
	M_3	58	58	85 %	82 %	79 %
	M_7	66	54,75	89 %	86 %	73 %

Tabuľka 2. Pravdepodobnosť že daná inštrukcia je relevantná pre daný počet funkcií z M_4

mutácia \ funkcie	aspoň dve	aspoň tri	všetky
0,025	57%	26%	22%
0,0025	52%	36%	12%
0,00025	58%	36%	24%

Tabuľka 3. Recyklácia inštrukcií vzhľadom k pôvodnému alebo skúmanému genómu.

mutácia	prostredie	pôvodný genóm		nový genóm	
		pozícia	inštrukcia	pozícia	inštrukcia
0,025	M_3	65%	57%	44%	38%
	M_7	65%	48%	34%	28%
0,0025	M_3	72%	60%	55%	45%
	M_7	80%	68%	42%	35%
0,00025	M_3	43%	4%	27%	2%
	M_7	57%	9%	20%	3%

Pri recyklácii inštrukcií sa najprv zamerajme na výsledky pre veľkosť mutácie 0,025 a 0,0025. V tabuľke 2 je uvedené percento využitia pozície a zároveň inštrukcie v readaptovanom genóme. V prvom prípade ide pomer k pôvodnému genómu adaptovanému v M_4 a v druhom prípade je to časť genómu, ktorý analyzujeme. Vidíme, že pri readaptácii na nové prostredie sa využili až dve tretiny pôvodných pozícií. Je zrejmé, že genóm má tendenciu znovu využívať už získanú štruktúru. Pri veľkosti mutácie 0,00025 vidíme, že genóm využíva 50 percent pôvodných pozícií. Treba si ale uvedomiť, že dĺžka genómu začína byť silne limitujúca pre túto veľkosť mutácie. Ako bolo uvedené skôr, pri malej mutácii narastá počet relevantných inštrukcií. Ako vidíme na recyklácii inštrukcií na rovnakej pozícii, tu počet prudko poklesol. Dá sa teda predpokladať, že genóm v obmedzených priestorových podmienkach značne zmenil svoju štruktúru. Táto zmena bola pravdepodobne vynútená potrebou získania komplexnejších vlastností.

4.2 Sledovanie informačnej zložitosti genómu

Vlastnosti genómu budeme študovať jednak po dobehnutí celej simulácie ako aj v čase získania všetkých odmeňovaných vlastností dominantným genómom. Vedie nás k tomu potreba mať populáciu v ekvilibriu, aby sme mohli správne odhadnúť entropiu populácie. Dá sa očakávať určitý pokles zložitosti smerom ku koncu simulácie, samozrejme až po nadobudnutí všetkých vlastností, spôsobený optimalizáciou genómu. Vo výsledkoch budeme porovnávať metódu zameranú na populáciu s metódou, ktorú som navrhol. Nebudú nás ani tak zaujímať absolútne čísla, pretože metódy sa líšia svojím zameraním, ako rozdiely medzi rôznymi množinami populácií. V tabuľke 4 vidíme výsledok merania podľa metódy zameranej na populáciu a v tabuľke 5 výsledky podľa mnou navrhnutej metódy.

Tabuľka 4.

prostredie		0,025	0,0025	0,00025
M_4	**čas zisku vlastností**	55.18	90.22	99.64
	koniec simulácie	39.49	57.69	96.79
M_7	**čas zisku vlastností**	71.19	88.68	99.32
	koniec simulácie	59.38	78.37	79.63

Tabuľka 5.

prostredie		0,025	0,0025	0,00025
M_4	**čas zisku vlastností**	22	24	57.95
	koniec simulácie	21.07	22.63	50.68
M_7	**čas zisku vlastností**	45	50.51	67
	koniec simulácie	42.07	47.63	63.68

Z výsledkov môžeme odvodiť závislosť rastu zložitosti nielen od zložitosti prostredia, ale aj od veľkosti mutácie. Organizmy vystavené menšej mutácii získavajú vlastnosti zapísané v jednoduchšej (kratšej) forme ako je to pri nízkych mutáciách. Zároveň je viditeľný pokles zložitosti na konci simulácie, pričom je tento pokles evidentnejší pre metódu zameranej na populáciu. Tento prudší pokles sa dá zdôvodniť už spomínaným efektom únosu. V metóde, ktorú som navrhol, je tento efekt odstránený, a preto je zrejmé, že pokles komplexnosti pri dlhšom behu populácie v ekvilibriu je prirodzenou vlastnosťou evolúcie.

5 Záver

Pri skúmaní vlastností genómu sa ukázal zásadný problém jednoznačnej identifikácie častí zodpovedných za konkrétnu vlastnosť. Zistili sme, že hlavne za jednoduchšie vlastnosti môžu byť potencionálne zodpovedné viaceré časti genómu. Ďalším problémom môže byť izolované skúmanie vlastností. Medzi zaujímavé vlastnosti asexuálneho prostredia patrí vznik modularity. Znamená to, že mechanizmy podporujúce vznik nejakej vlastnosti môžu byť zastúpené v inom prostredí iným rozdielnym mechanizmom. Už pri malom počte schopností organizmu sa prejavila komplexnosť genómu. Komplexnosť sa prejavuje veľkým preťažovaním inštrukcií. Ak by sme chceli upraviť nejakú vlastnosť genómu bez zmeny ostatných, museli by sme vykonať rozsiahlejšie zmeny. Získané štruktúry v genóme však organizmus po zmene prostredia dokáže efektívne využiť. Dôležitou vlastnosťou organizmu je schopnosť rýchlo sa readaptovať v premenlivom prostredí. Schopnosť recyklovať štruktúry môže byť rozhodujúca. V takom prípade by bolo zaujímavé sledovať viacnásobné zmeny v prostredí, ktoré by mohli byť dôvodom na ďalší nárast komplexnosti.

Preukázali sme súvis rastu zložitosti nielen v závislosti od prostredia, ale aj od veľkosti mutácie. Je preto dôležité pri porovnávaní organizmov brať do úvahy aj tento faktor, nakoľko zdanlivo komplexnejšia povaha daného genómu môže byť spôsobená iba inou veľkosťou mutácie. Môže sa teda stať, že v až na mutáciu totožných prostrediach, sa

jednoduchší organizmus (z pohľadu množstva fenotypických vlastností) bude javiť ako komplexnejší a pri zmene veľkosti mutácie získame opačné výsledky.

Poďakovanie: Tento príspevok vznikol za podpory grantovej agentúry VEGA SR v rámci grantovej úlohy 1/4053/07 a za podpory grantovej agentúry APVT v rámci grantovej úlohy APVT-20-002504.

Literatúra

[1] Huang, W., Ofria, C., Torng, E.: *Measuring Biological Complexity in Digital Organisms*, Proceedings of the Ninth International Conference on Artificial Life, Boston MA, Sept 12-15, 2004: 315-321.
[2] Adami, C.: Sequence complexity in Darwinian evolution. *Complexity* **8** (2003) 49-56.
[3] Bennett C. H.: *Logical depth and physical complexity*. The Universal Turing Machine, A Half-Century Survey. Oxford: Oxford University Press 1988, pp. 227-257.
[4] Lenski, R. E., C. Ofria, R. T. Pennock, and C. Adami: The evolutionary origin of complex features. *Nature* **423** (2003) 139-144.
[5] Wilke, C. O., Wang, J. L., Ofria, C., Lenski, R. E., Adami, C.: Evolution of digital organisms at high mutation rates leads to survival of the flattest. *Nature* **412** (2001) 331-333.
[6] Adami, C., Cerf, N.J.: Physical complexity of symbolic sequences., Physica D **137** (2000) 62-69.
[7] Wagenaar, D. A., Adami, C.: Influence of Chance, History, and Adaptation on Digital Evolution. *Artificial Life* **10** (2004) 181-190.
[8] Kvasnička, V., Pospíchal, J., Tiňo, P.: *Evolučné algoritmy*. Bratislava: Vydavateľstvo STU 2000.

Lindenmayerove systémy a ich tvorca v spomienkach, faktoch a citátoch

Alica Kelemenová[1]

Abstrakt. Príspevok – aj na základe osobných spomienok autorky – približuje vznik, rozvoj a niektoré výsledky teórie Lindenmayerových (L) systémov.[2]

Aristid Lindenmayer
prednášajúci v rámci IMYCS´86 na Smolenickom zámku v októbri 1986
svoju tutoriálnu prednášku [38]

[1] Ústav informatiky, Slezská univerzita, Bezručovo nám. 13, 74601 Opava, Česká republika a Katedra informatiky, Katolícka univerzita, Nám. A. Hlinku 56, 03401 Ružomberok, Slovensko,
E-mail: kelemenova@fpf.slu.cz
[2] Prednáška odznela na KUŽ V pri príležitosti nedožitých 80-tin A. Lindenmayera (1925-1989)

1 Moje prvé stretnutie so systémami A. Lindenmayera

Písal sa rok 1973 a po roku pôsobenia na Matematickom ústave SAV v Bratislave som dostala prvú príležitosť zúčastniť sa hneď dvoch medzinárodných vedeckých podujatí: letnej školy z teórie pologrúp, ktorú usporiadala Univerzita Attilu Józsefa v maďarskom Szegede, a v náväznosti na ňu 2. medzinárodného sympózia o matematických základoch teoretickej informatiky (Mathematical Foundations of Computer Science, MFCS´73), ktorú pod záštitou Matematického ústavu SAV ogranizoval kolektív pod vedením Jozefa Grusku na Štrbskom Plese vo Vysokých Tatrách. Arto Salomaa vo svojich prednáškach na týchto podujatiach s neskrývaným nadšením predstavil vývojové systémy, ktoré v tých rokoch zaviedol Aristid Lindenmayer. Hovoril o nich ako o jednom z hlavných trendov výskumu oblasti teórie automatov a formálnych jazykov tých čias [51].

Model navrhnutý Lindenmayerom poskytoval možnosť zapísať jednoduchý vláknitý organizmus chápaný ako reťazec buniek pomocou reťazca symbolov (písmen) a zachytiť vývoj takéhoto organizmu pravidlami určujúcimi vývoj jednotlivých buniek. Vývoj celého organizmu vznikal ako dôsledok používania pravidiel na všetky bunky reťazca súčasne. Reťazce symbolov sa pretvárali na základe pravidiel podobných pravidlám, s ktorými sme sa stretali v oblasti klasickej teórie formálnych gramatík. Vznikol nový variant generatívnych gramatík, (ktoré sa postupne začali označovať ako L systémy) charakterizovaná abecedou, prepisovacími pravidlami, reťazcom symbolov predstavujúcim počiatočné vývojové štádium modelovaného organizmu a paralelným odvodzovaním. Typická L systémová „vetvička“ znázorňujúca vývoj riasy sa dostala aj na obal práve vtedy vychádzajúcej monografie Arto Salomaa [52].

Základné odlišnosti novozavedených vývojových systémov v porovnaní s klasickými gramatikami Chomského typu (Hopcroft, Ullman 1969) vyplývajú z ich biologickej motivácie. Okrem už spomínaného aplikovania pravidiel súčasne na všetky symboly je pre tento systém charakteristické, že všetky symboly reťazcov reprezentujú aktuálne stavy jednotlivých buniek, systém pri generovaní reťazcov nepoužíva tzv. pomocné, rýdzo pracovné symboly. Absolútny paralelizmus prepisu symbolov a generovanie reťazcov bez pomocných symbolov, ktorým dovtedy teória formálnych jazykov a gramatík len okrajovo venovala pozornosť, fascinovali odborníkov a poskytli stimuly pre nebývalo intenzívny rozvoj a formovanie teórie L systémov.

Výskum L systémov od ich zrodu prakticky až podnes prebieha po niekoľkých vzájomne voľne prepojených líniách. Okrem už spomínanej príslušnosti do teórie formálnych jazykov a gramatík sa L systémy študujú aj v rámci teoretickej biológie, významne ovplyvnili rozvoj počítačovej grafiky a je to aj oblasť, ktorá zreteľne predznamenala vznik novej disciplíny, umelého života [35].

V oblasti počítačovej grafiky nachádzame L systémy ako bázu pre vizualizáciu a sčasti aj animáciu vetvenia rastlín, pre vykresľovanie tvaru listov alebo čiarových obrazcov na ulitách mäkkýšov. Túto oblasť rozvoja L systémov reprezentuje azda najprimeranejšie knihy [46,19,41] obsahujúce množstvo niekedy veľmi prekvapivých a esteticky veľmi pôsobivých ilustrácií (umelej) rastlinnej a podmorskej ríše spolu s návodmi, ako ich vytvárať. Jedná sa o vizuálne pôsobivú a veľmi vďačnú tému, ktorú nájdeme často spracovanú na www stránkach. Súčasný stav verne reflektujú osobné www

stránky P. Prusikiewicza. Na konferenciách Kognice a umělý život bol tento prístup prezentovaný v prácach [17,18].[3]

V ďalšej časti príspevku sa podrobnejšie zameriame na okolnosti, ktoré podnietili A. Lindenmayera k výskumu systémov, ktoré boli neskôr podľa neho pomenované, ako výskum L systémov obohatil teóriu formálnych jazykov a čím prispeli prostriedky teórie formálnych jazykov ku spoznávaniu L systémov.

2 Lindenmayerova cesta k vývojovým systémom

Aristid Lindenmayer sa narodil 17. novembra 1925 v Budapešti, v židovskej rodine patriacej do nižšej strednej vrstvy vtedajšej maďarskej mestskej spoločnosti. Maturoval na preslávenom (a dodnes fungujúcom) Evanjelickom gymnáziu na Alejovej ulici (Fasor utca) v Budapešti, kde maturovali – ako uvádza [22] – aj také osobnosti svetovej vedy, ako obdivuhodne všestranný John (pôvodne János) von Neumann, fyzici E. Wiegner a otec vodíkovej bomby E. Teller, alebo významní matematici 20. storočia G. Pólya, T. Kármán a P. Erdös.

O jeho ďalších osudoch však nechajme hovoriť samotného Lindenmayera[4]:

Prešiel som vskutku dosť dlhú cestu. O problematiku pôvodu biologických štruktúr (morfogenézu) som sa začal zaujímať už v začiatkoch svojich štúdií. Profesor Buzágh bol ten, kto ma počas mojich štúdií fyzikálnej a koloidnej chémie na univerzite v Budapešti upozornil na potrebu teoreticky vysvetliť tie druhy nespojitostí, ktoré pozorujeme v koloidných suspenziách a v géloch. Na prednáškach z biológie a z chémie som na tento problém narazil opäť: Akým spôsobom je možné opísať štruktúry, ktoré sú v priestore alebo v čase čiastočne diskrétne a čiastočne spojité? Šlo mi o formulácie v abstraktnom jazyku odlišnom od spojitej matematiky a ja som sa domnieval, že by sme ich mohli vytvoriť v rámci matematickej logiky. Po dlhom hľadaní som roku 1952 narazil na referenciu Woodgerovej práce (v tom čase som bol študentom (graduate student) biológie na Michiganskej univerzite v Ann Arbor). Woodger sa približne od roku 1930 pokúšal formalizovať ústredné biologické teórie, medzi nimi aj genetiku a vývojovú biológiu. Bol to on, kto ako prvý aplikoval matematickú logiku v biológii v duchu Whiteheadovho a Russellovho Principia Mathematica. *Zastával názor, že relačné a množinovo-teoretické pojmy sa hodia na vyjadrenie mnohých biologických fenoménov väčšmi než analytické. Napriek tomu, že jeho formalizácie vývoja, dedičnosti a evolúcie nezískali popularitu medzi biológmi, najmä preto, že z filozofických dôvodov zdôrazňoval pozorovania a stránil sa teoretických koncepcií, jeho základné východiská zdôrazňujúce vhodnosť relačného, nekvantitatívneho a neanalytického formalizmu pre biológiu sa značne rozšírili koncom 50. rokov.*

Po získaní doktorátu filozofie (PhD.) na Michiganskej univerzite som strávil akademický rok u Woodgera v Londýne v Anglicku. Učil som sa od neho logiku a vypracovával som axiomatický systém na opis životných cyklov. Už vtedy som sa zaoberal problémami rastu a tvarov. Práve tieto problémy upriamili moju pozornosť na funkciu buniek a ich vzájomné vzťahy v priebehu vývoja viacbunkových organizmov, teda na to, čo vo svojej axiomatizácii zdôrazňoval aj Woodger. V tom období až do roku 1968 som sa zaoberal vo Philadelphii a v New Yorku aj experimentálnou prácou. Po takýchto

[3] Ako prvá stručná informácia o tomto prístupe poslúži aj kapitola v [3].
[4] Úryvok je z rozhovoru publikovaného v anglickej verzii ako [28] a v slovenskej ako [29].

skúsenostiach nadišla chvíľa môjho prvého stretnutia s teóriou automatov. Bolo to v rokoch 1962 - 63, keď som pracoval v biomatematickej skupine Univerzity štátu Severná Karolína, kde som mal často možnosť stretávať sa s embryológom J. R. Greggom z Dukovej univerzity. Aj on predtým spolupracoval s Woodgerom a napísal články o množinovo-teoretických základoch taxonómie a embryológie. V tých časoch prišiel do kontaktu s teóriou automatov (ktorá bola nastolená v r. 1956 prácami Moora a Mealyho) a spoločne sme študovali Ginsburgov prvý systematický výklad tejto teórie[5]*. Približne v tom istom čase vyšla Ashbyho kniha*[6]*, ktorá sprístupnila teóriu riadenia aj pre nematematikov (už v roku 1963 som používal túto knihu pri prednáškach pre biológov). Zbehlosť v logike, ktorú sme obaja získali u Woodgera, bola pre nás, biológov nevyhnutná k tomu, aby sme v tom čase mohli preniknúť do teórie automatov.*

3 Pôvodný model

L systémy, ktoré sú prezentované ako formálne gramatiky už v prácach z r. 1971, boli pôvodne zavedené (samozrejme nie pod názvom L systémy) ako polia konečných automatov. Na obdobie formovania modelu spomínal v roku 1984 A. Lindenmayer, vtedy profesor teoretickej biológie na univerzite v Utrechte takto[7]:

Vznik L systémov sa viaže k jednému problému z Ashbyho knihy[5]*. V kapitole o konečných automatoch (cvičenie 4.7.7) sa pýta: Aké chovanie je možné očakávať od dlhého reťazca obojsmerne pospájaných konečných automatov v prípade, že tieto dostávajú vstupy od oboch susedných automatov? Samozrejme, v Kleeneho práci*[8] *je možné nájsť odpoveď pre prípad siete konečných automatov konštantného rozsahu: Chovanie takéhoto reťazca je v zásade rovnaké ako chovanie jedného konečného automatu. Mňa však znepokojovala otázka chovania rozrastajúceho sa poľa konečných automatov. Takéto polia totiž môžu poslúžiť ako realistická simulácia rastu bunkových vlákien aké nachádzame napríklad u rias, machov či húb. Riešenie tohto problému sa však ukázalo byť náročnejšie, než som očakával. To ale nič nemení na fakte, že v roku 1968 boli definované L systémy ako interagujúce rozrastajúce sa lineárne polia konečných automatov.*

V pôvodnej práci [36], zaslanej do tlače v auguste 1967, v ktorej bol model navrhnutý ešte v terminológii teórie automatov, čítame:

Cieľom práce je navrhnúť teoretickú bázu, v rámci ktorej môžeme študovať, počítať a porovnávať medzibunkové vzťahy. Chceme sa zamerať výlučne na tie aspekty vývoja, ktoré sú dôsledkom interakcií dvoch alebo viacerých buniek, časovania a umiestnenia delenia buniek a indukcie medzi bunkami .... Pokúsil som sa vytvoriť matematickú teóriu, v rámci ktorej môžeme vysloviť a overiť vety a môžeme rigorózne odvodiť a porovnať dôsledky presne stanovených predpokladov.

---

[5] Ginsburg, S.: *An introduction to Mathematical Machine Theory*. Addison-Wesley, Reading, Mass., 1962.
[6] Ashby, W. R.: *An Introduction to Cybernetics*. Chapman and Hall, London, 1961.
[7] Úryvok je z rovnakého rozhovoru, ako predchádzajúci.
[8] Kleene, S. C: Representation of events in nerve nets and finite automata. In: Automata Studies (C. E. Shannon and J. McCarthy, eds.) Princeton University Press, Princeton, NJ, 1956, 1-41.

Požiadavky, z ktorých navrhovaný model vychádza sú nasledovné:

1. *Študujeme bunkové organizmy. Každú bunku považujeme za samostatnú jednotku s priestorovým a časovým ohraničením.*
2. *Budeme pracovať s diskrétnym časom, t.j. časom, ktorý postupuje v oddelených intervaloch (tieto môžu byť tak krátke ako je potrebné).*
3. *Bunka sa v každom časovom okamihu môže nachádzať v jednom z určitého počtu stavov, ktoré môžeme interpretovať ako určité rozpoznateľné fyziologické alebo morfologické podmienky, napríklad prítomnosť určitých enzýmov alebo sub-celulárnych štruktúr, alebo polohu a rozmer bunkových stien.*
4. *Bunka v každom časovom okamihu vytvára jeden výstup z danej množiny výstupov, ktoré môžeme interpretovať ako tvorbu rozptýliteľnej chemickej látky alebo elektrických impulzov, alebo ľubovoľného iného typu fyzikálnych či chemických zmien, ktoré sa môžu šíriť mimo bunky. Ak bunka vytvorí viac takýchto zmien, potom ich súbor slúži ako jednoduchý výstup.*
5. *Bunka môže v každom časovom okamihu dostať impulzy od svojich susedov, ich súbor vytvára vstup pre bunku v danom okamihu.*
6. *Nasledujúci stav bunky je v každom okamihu funkciou jej aktuálneho stavu a vstupu (δ funkcia).*
7. *Taktiež výstup bunky je v každom okamihu funkciou jej aktuálneho stavu a vstupu (λ funkcia).*
8. *Je potrebné stanoviť explicitné pravidlá ako sa majú konvertovať výstupy na vstupy.*
9. *Je potrebné určiť susedné bunky, t.j. určiť výstupy ktorých buniek budú vytvárať vstupy danej bunke.*
10. *Je potrebné brať do úvahy delenie buniek, preto hodnotou funkcie určujúcej nasledujúci stav môže byť pre niektoré kombinácie stav – vstup postupnosť viacerých stavov. Takto môže byť napríklad $\delta(a,1) = ab$ pre stavy a, b a vstup 1.*
11. *Každá bunka zachováva svoju identitu od svojho vzniku, až kým sa rozdelí alebo nezanikne. Dcérine bunky, ktoré vzniknú delením, považujeme za neidentické s bunkou matky.*
12. *Zánik bunky je možné reprezentovať zavedením prázdneho stavu e, ktorý v každom prípade potrebujeme z teoretických dôvodov, máme teda $\delta(a,1) = e$, pre určité kombinácie stavov a vstupov.*

Model, vytvorený na základe hore vymenovaných podmienok je zostavený z konečných automatov tak, že vstupy a výstupy automatov sú identické s ich stavmi. Každý automat má teda rovnakú množinu stavov, množinu vstupných symbolov a množinu výstupných symbolov. Výstup automatu je určený priamo jeho stavom. V práci sú uvedené modely pre vláknitý organizmus s jednostranným (konkrétne ľavostranným) vstupom, modely pre vláknitý organizmus s obojstranným vstupom a modely pre rozvetvený organizmus. V prvom z uvedených modelov, neskôr označovanom 1L systém, je výstupom automatu, nachádzajúceho sa v stave p nezávisle na jeho vstupe q hodnota p, platí teda $\lambda(p,q) = p$. Prechodová funkcia $\delta(p,q)$ môže v tomto modeli priradiť stavu automatu p a vstupu q reťazec stavov w, $\delta(p,q) = w$ (v prípade, že reťazec má aspoň dva prvky, predstavujú jednotlivé prvky reťazca stavy nových identických automatov vložených za sebou na mieste pôvodného automatu). V druhom z uvedených modelov,

neskôr označovanom 2L systém, pracujeme s prechodovou funkciou δ a s ľavou λ_L a pravou λ_R výstupnou funkciou závislou na stave *p,* ľavom vstupe automatu *q* a pravom vstupe automatu *r*. Prechodová funkcia priraďuje reťazec stavov $\delta(p,q,r) = w$, pre hodnoty oboch funkcií výskytu platí $\lambda_L(p,q,r) = \lambda_R(p,q,r) = p$. Simulácia rastu rozvetveného organizmu je založená na reprezentácii vetvenia pomocou zátvoriek [,]. Reťazec symbolov umiestnený v zátvorke tvorí vetvu nad bunkou napísanou bezprostredne pred symbolom [. Symboly vetvenia [,] sa pri prepisovaní nemenia, môžu však vznikať. Modely rozvetvených rastlín (štruktúr) v pôvodnej práci si zaslúžia pozornosť hneď z dvoch príčin. Prvýkrát tu autor zmieňuje aj model pracujúci nezávisle na vstupných údajoch od susedných automatov, neskôr nazývaný 0L systém. Je tu vlastne prvýkrát diskutovaný problém ekvivalencie rastu a vetvenia. Príklad dvoch rozličných modelov, ktoré vytvárajú rovnakú vetviacu štruktúru slúži na porovnanie ekvivalentných systémov z hľadiska ich interakcií, počtu symbolov a tvaru prechodových funkcií spôsobujúcich predlžovanie štruktúry.

Veľmi špeciálny tvar prechodovej funkcie navrhnutého modelu, ako aj skutočnosť, že všetky tri množiny: množina stavov, vstupov a výstupov sú identické, umožnili zapísať navrhnuté systémy jednoduchšie, ako formálnu gramatiku $G = (\Sigma, P, w_0)$ s jednou abecedou Σ, s pravidlami z P pričom pravidlo v tvare $(q, p, r) \rightarrow w$ nahradilo prechodovú funkciu $\delta(p, q, r) = w$, kde *q* a *r* môžu byť aj reťazce symbolov a so začiatočným reťazcom w_0, nazývaným tiež axióma systému. Hovoríme o interaktívnych L systémoch, skrátene IL systémoch. V prípade, že sú dĺžky reťazcov *q* a *r* ohraničené po rade číslami *m*, *n* spresňujeme, že sa jedná o (*m*, *n*)L systém. Systém pracuje tak, že zo začiatočného reťazca súčasným prepisovaním všetkých písmen reťazca na základe daných pravidiel vytvára stále nové reťazce.

Aristid Lindenmayer veľmi starostlivo sledoval rozvoj teórie, ktorej zrod inicioval. Na konferenciách a letných školách ho bolo možné stretnúť s obalom plným priesvitiek, často a rád zasväcoval záujemcov do tajov vývojových systémov. Boli sme radi, že sme ho mohli dvakrát privítať aj na Slovensku. V roku 1983 bol hosťom na Matematickom ústave SAV. Vyznamenaním pre nás bolo, že profesor Lindenmayer prijal naše pozvanie a vystúpil s tutoriálnou prednáškou na štvrtom medzinárodnom stretnutí mladých informatikov, ktoré sme usporiadali roku 1986 v priestoroch Domova vedeckých pracovníkov SAV na zámku v Smoleniciach.

4 Rast a rastové funkcie

Lindenmayer neustále preferoval deterministické systémy, ktoré umožňujú z daného reťazca vytvoriť nasledujúci reťazec jediným možným spôsobom. Zastával názor, že prirodzenú variabilnosť vývoja rastliny je možné zachytiť priradením parametrov k jednoznačným (deterministickým) vývojovým pravidlám systému. Deterministické systémy umožnili prehľadne formulovať prirodzené požiadavky na spoznávanie kvality a kvantity rastu. Boli preto východiskom pre zavedenie a výskum rastových funkcií L systémov, teda tých funkcií, ktoré určujú dĺžku vygenerovaného reťazca v závislosti na počte vývojových štádií, ktoré predchádzali vytvoreniu reťazca.

Široká škála dnes známych poznatkov o rastových funkciách sa týka najmä nasledujúcich okruhov: súvislosti typu rastovej funkcie a štruktúry (pravidiel) systému; problému ekvivalencie rastových funkcií L systémov, t.j. stanovenia podmienok, za

ktorých majú rôzne systémy rovnakú rastovú funkciu; otázok rozhodnuteľnosti a zložitosti stanovenia typu rastových funkcií; metód syntézy a analýzy rastových funkcií pre rôzne typy systémov. K formulácii a na získanie týchto výsledkov sa využívajú dobre rozpracované matematické disciplíny, napr. teória matíc alebo mocninných radov.

Užitočným čítaním o rastových funkciách, je sviežim štýlom písaný prehľad [48], ktorý sprostredkuje dosiahnuté výsledky s minimom formalizmu. V jej úvode čítame:

Štúdium rastových funkcií vyžaduje podstatne hlbšie znalosti z matematiky než ostatné oblasti teórie L systémov. ... Dúfame, že nasledujúce vysvetlenie neodradí čitateľov bez hlbších znalostí matematiky, a že tento prístup otvorí nové perspektívy i pre matematicky zdatných čitateľov. Pokúsime sa však vysvetliť teóriu rastových funkcií s čo najmenším počtom vzorcov.

Z rozhovoru vo vlaku medzi odborníkom na formálne jazyky, matematikom a cestovateľom – skúseným laikom, sa o rastových funkciách postupne dozvedáme napríklad, že:

Dĺžku *i*-teho odvodeného reťazca deterministického L systému môžeme zhora ohraničiť číslom c^{i+1}, kde konštanta *c* je daná maximom dĺžok pravých strán pravidiel L systému. Najrýchlejší rast L systému teda nemôže byť rýchlejší než exponenciálny. Pri skúmaní rastu systémov bez interakcie, 0L systémov s pravidlami v tvare $p \rightarrow w$, kde *p* je symbol, ktorý sa má zameniť reťazcom *w* nezáleží na poradí symbolov v reťazci. Sledujeme len počet výskytov jednotlivých symbolov v reťazci, čo zaznamenáme v tvare *Parikhovho* vektora reťazca. Na určenie dĺžky nasledujúceho reťazca potrebujeme poznať tzv. *Parikhov vektor* reťazca, znalosť celkovej dĺžky reťazca nestačí. Rôzne symboly môžu totiž rásť veľmi odlišne. Všetky potrebné informácie o raste D0L systému je možné uchovať v matici rastu *M*. Riadky a stĺpce matice sú zviazané s výskytom písmen presne podľa zvoleného usporiadania. Napríklad tretí prvok v druhom riadku určuje, koľkokrát sa tretie písmeno vyskytuje na pravej strane pravidla pre druhé písmeno, t.j. koľko výskytov tretieho písmena nájdeme v prepise druhého písmena. Pokiaľ sa zaujímame len o dĺžku vytvoreného slova, tak je možné vyjadriť počet vytvorených symbolov na každom kroku odvodenia umocnením matice rastu. Počet jednotlivých symbolov reťazca po *n* krokoch odvodenia je daný *Parikhovym* vektorom πM^n. Hodnotu *f(n)* rastovej funkcie dostaneme súčtom jednotlivých zložiek Parikhovho vektora, čo umožňuje jeho vynásobenie s *n*-rozmerným jednotkovým stĺpcovým vektorom η. Rastovú funkciu môžeme teda vyjadriť v tvare $f(n) = \pi M^n \eta$, kde riadky matice určujú násobnosť výskytu písmena v pravidlách a zložky vektora π určujú násobnosť výskytu jednotlivých písmen v axióme.

Toto vyjadrenie umožnilo pomocou teórie matíc odvodiť rad základných poznatkov o rastových funkciách. Z nich môžeme uviesť napríklad rekurzívny vzťah určujúci hodnotu rastovej funkcie na základe jej hodnôt pre nižšie argumenty

$$f(n+i) = a_1 f(n + i - 1) + \ldots + a_n f(n) \text{ pre } n \geq 0$$

(odvodený použitím Cayley-Hamiltonovej vety o charakteristickej rovnici matice).

Informácia uchovaná v matici rastu a získaný rekurzívny vzorec dovoľujú využiť množstvo poznatkov z teórie matíc a diferenčných rovníc. Hodnoty *f(n)* môžeme napríklad zapísať pomocou súčinu exponenciálnej funkcie μ^n a polynómu premennej *n*, teda ako konečný súčet členov v tvare:

$$(\beta_0 + \beta_1 n + \ldots + \beta_t n^t)\, \mu^n,$$

kde μ je koreň charakteristickej rovnice matice M násobnosti t. Čísla μ a $\beta_0, \ldots, \beta_t$ v tomto vyjadrení môžu byť komplexné, aj keď sú hodnoty $f(n)$ nezáporné.

Rastová funkcia D0L systému (deterministického systému bez interakcií) môže byť konštantná, polynomiálna alebo súčet členov, ktoré sú súčinom polynómu a exponenciálnej funkcie, ako to ukazuje predchádzajúci vzťah. U interaktívnych systémov sa stretávame s bohatšou škálou možností. Príklady DIL systémov s logaritmickým rastom a subexponenciálnym rastom nájde čitateľ v [20,48].

Ďalším zaujímavým výsledkom je lema o dlhých konštantných intervaloch. Predstavme si, že f je rastová funkcia *D0L* systému a pre každú konštantu k existuje n také, že k po sebe vygenerovaných slov počnúc n-tým má rovnakú dĺžku, teda formálne

$$f(n) = f(n+1) = \ldots = f(n+k)$$

Potom f je konečná a ohraničená, t.j. existuje n_0 také, že $f(n) = f(n_0)$ pre všetky $n \geq n_0$.

Maticová reprezentácia poskytuje jednoduchú metódu ako rozhodnúť o rastovej ekvivalencii D0L systémov. Dva D0L systémy s n prvkovou abecedou a n' prvkovou abecedou majú rovnakú rastovú funkciu pokiaľ sa rovnajú dĺžky prvých $n + n'$ reťazcov odvodených v obidvoch systémoch. Toto je najlepšia možná hranica [44]. Uvedieme príklad dvoch rozličných D0L systémov s rovnakou rastovou funkciou. Systémy majú rovnakú axiómu. Prvý systém sa riadi pravidlami $a \to ab^3$, $b \to b^3$ a druhý $a \to acde$, $b \to cde$, $c \to b^2d^2$, $d \to d^3$, $e \to bd$.

Tvar rastovej funkcie $f(n) = \pi M^n \eta$ podnietil výskum zovšeobecnených funkcií tohto tvaru aj pre ľubovoľné hodnoty (celočíselné a racionálne) vyskytujúce sa ako zložky vektorov π a η a matice M. Množstvo výsledkov o tomto probléme sa nájde v monografii [54].

Rastovým funkciám a funkciám výskytu D0L systému vo vzťahu k zložitosti jeho zápisu som sa venovala na Matematickom ústave SAV v Bratislave [23,24]. Vlastnosti podielu rastových funkcií skúmala autorka prvej kandidátskej dizertačnej práce na Slovensku, zameranej na problematiku L systémov Mária Králová [33,34].

5 Problém D0L ekvivalencie, jeho riešenie a niektoré dôsledky

V predchádzajúcej kapitole sme uviedli, že problém či dva D0L systémy majú rovnakú rastovú funkciu je rozhodnuteľný. Všeobecnejšej otázke, otázke D0L ekvivalencie, sa budeme venovať v tejto kapitole. Tento problém formuloval začiatkom sedemdesiatych rokov minulého storočia Lindenmayer pre systémy bez vymazávania [37] a je to problém, ktorý napriek intenzívnej pozornosti bádateľov odolával vyriešeniu až do roku 1976 [7]. Úloha zistiť, či je možné algoritmicky rozhodnúť, že sú slová vygenerované v dvoch daných D0L systémoch ekvivalentné, je jedným z najznámejších problémov teórie L systémov. Nielen samotná jednoduchá formulácia problému, ale aj techniky vyvinuté na jeho pokorenie a nové okruhy problémov vyvolané snahou o jeho vyriešenie poznamenali výskum na dlhé roky aj po uverejnení prvého algoritmu rozhodujúceho ekvivalenciu D0L systémov. Zhrnieme preto, ako postupovalo riešenie tohto problému.

Problém ekvivalencie nachádzame prvýkrát formulovaný v januári 1972 ako otvorený problém pre 0L systémy (nedeterministické systémy bez interakcie) a pre P0L systémy v [37]. V roku 1973 A. Salomaa dokázal nerozhodnuteľnosť tohto problému pre

0L systémy. Súčasne sa objavuje hypotéza, že pre D0L systémy je tento problém rozhodnuteľný [52].

Kompletnému riešeniu problému predchádzali čiastočné výsledky o rozhodnuteľnosti špeciálnych prípadov:

- ***1973*** – ekvivalencia rastových funkcií [44]
- ***1974*** – ekvivalencia Parikhových vektorov[9] [43]. Významný posun v riešení predstavuje tvrdenie uverejnené v tej istej práci, že D0L ekvivalencia jazykov je rozhodnuteľná práve vtedy, ak je rozhodnuteľná ekvivalencia D0L postupností. Ako ukazuje nasledujúci príklad slová v ekvivalentných D0L systémoch môžu byť vytvárané v odlišnom poradí:

  $$G_1 = (\{a, b\}, \{a \to b^2, b \to a\}, b) \quad G_2 = (\{a, b\}, \{a \to b, b \to a^2\}, a).$$

  Prvý systém generuje slová jazyka $L = \{a^i: i = 2^j, j \geq 0\} \cup \{b^i: i = 2^j, j \geq 0\}$ v poradí

  $$b \Rightarrow a \Rightarrow bb \Rightarrow aa \Rightarrow bbbb \Rightarrow aaaa \Rightarrow \ldots$$

  a druhý

  $$a \Rightarrow b \Rightarrow aa \Rightarrow bb \Rightarrow aaaa \Rightarrow bbbb \Rightarrow \ldots$$

- ***1975***[10] – ekvivalencia jednoduchých[11] D0L systémov [6].
  ekvivalencia D0L systémov nad dvojprvkovou abecedou [55][12];
- ***1976*** – ekvivalencia pre polynomiálne ohraničené D0L systémy [21].
  K. Čulík II a I. Friš oznamujú rozhodnuteľnosť problému D0L ekvivalencie [7]

Prvý publikovaný algoritmus ako aj algoritmus uverejnený v [10] sú veľmi náročné. Pracujú s kódmi s ohraničeným zdržaním, s elementárnymi morfizmami, zavádza sa jazyk rovnosti dvoch morfizmov a študuje jeho regulárnosť. Je preto prirodzené, že záujem o tento problém pretrváva naďalej a sústreďuje sa na hľadanie jednoduchšieho riešenia. Objavujú sa nové dôkazy už známeho tvrdenia, nové techniky a prístupy k riešeniu a tým aj nové okruhy súvisiacich problémov [9,50,13]. Asi najvýraznejšie súčasné trendy sú charakterizované v [14], keď v súvislosti so skúmaným problémom a jeho zovšeobecneniami hovorí o voľných monoidoch morfizmov a ich iteráciách.

Vráťme sa k hľadaniu jednoduchého algoritmu. Výzvou je dodnes otvorený problém zistiť, podobne ako pri ekvivalencii rastových funkcií, koľko dvojíc generovaných slov je potrebné porovnať, aby sme rozhodli o rovnosti D0L postupností. Analýzy známych dôkazov poskytujú enormne veľké číslo. V tejto súvislosti je očakávaný výsledok známy pod názvom *Hypotéza 2n.*

Hypotéza 2n: Ak sa dve D0L postupnosti nad *n* prvkovou abecedou zhodujú v prvých *2n* členoch, potom sú ekvivalentné.

Príkladom dvojice D0L systémov, ktoré ukazujú, že hodnotu *2n* v hypotéze nie je možné znížiť, je dvojica G a H nad dvojprvkovou abecedou, s axiómou *ab* a pravidlami

$$G: a \to abb,\ b \to aabba \qquad H: a \to abbaabb,\ b \to a$$

---

[9] Parikhov vektor slova udáva počet výskytov jednotlivých písmen v tomto slove.
[10] Uvádzame rok(y), kedy bol výsledok dokázaný. Časopisecká publikácia mohla byť zverejnená neskôr.
[11] L systém je jednoduchý ak pre niektoré *k* sú všetky prvky M^k *k*-tej mocniny rastovej matice nenulové.
[12] Prezentovaný dôkaz sa nedá zovšeobecniť na viacprvkovú abecedu.

Tieto D0L systémy generujú v prvých troch krokoch odvodenia rovnaké slová.

$$G: ab \Rightarrow abb\ aabba \Rightarrow abb\ aabba\ \underline{a\ a}bba\ abb\ abb\ aabba\ aabba\ abb \Rightarrow w\ abba\mathbf{b}bu$$
$$H: ab \Rightarrow abbaabb\ a \Rightarrow abb\ aabba\ \underline{aa}\ bb\ aabb\ abb\ aabbaa\ a\ bbaabb \Rightarrow w\ abba\mathbf{a}bbv$$

Tretie odvodené slovo obsahuje ako podreťazec *aaa* (jeho posledné dve písmená sme v odvodení zvýraznili) čo zapríčiní, že každý z uvedených systémov vytvorí vo štvrtom kroku iný reťazec (pre prehľadnosť sme rovnako sa prepisujúcu časť na začiatku slova nerozpisovali a uviedli len miesto, kde sa budú slová odlišovať, prvé rozdielne písmeno je zvýraznené.) Príklad je možné zovšeobecniť aj na viacprvkovú abecedu.

6 Regenerácia v L systémoch

Počas mojich dvoch pobytov u A. Lindenmayera v holandskom Utrechte, vo februári 1984 a v marci 1988 sme často diskutovali o regenerácii a o možnostiach, ktoré poskytujú L systémy pre štúdium tejto schopnosti živých organizmov.

Skôr ako sa zamyslíme nad tým, ktoré vlastnosti L systému umožňujú jeho schopnosť regenerácie a za akých podmienok, uvedieme príklad systému s touto vlastnosťou. Vzhľadom na to, že systém musí dokázať rozpoznať, že niekde v jeho reťazci došlo k porušeniu vývoja, je samozrejmé, že systém schopný regenerácie bude systém s interakciou. Ako ukážku L systému s úplnou regeneráciou (všetkých neprázdnych podreťazcov) uvedieme (1,1) L systém pre regeneráciu reťazca ab^4c. Navrhnutý systém umožňuje ľubovoľnému podreťazcu daného reťazca, aby dorástol na pôvodný reťazec. Tento sa už ďalej nevyvíja, zostáva rovnaký.

Pravidlá systému sú zapísané v nasledujúcich tabuľkách. Symbol v ľavom hornom rohu tabuľky určuje prepisovaný symbol, v prvom stĺpci príslušnej tabuľky je symbol, ktorý sa nachádza pred prepisovaným symbolom a v prvom riadku je symbol, ktorý nasleduje za prepisovaným symbolom. Reťazec, na ktorý sa symbol prepisuje, je uvedený v príslušnom políčku tabuľky. Napríklad $(-, a, c) \rightarrow ab$, $(a, b, a) \rightarrow bb$.

Pravidlo $(a, b, a) \rightarrow bb$ sa použije na prepis písmena *b* uvedeného v strede trojice v situácii, keď pred ním aj za ním stojí písmeno *a*, na reťazec *bb* uvedený za šípkou. Čiarka – reprezentuje situáciu, keď pred alebo za prepisovaným písmenom nestojí žiadne písmeno, teda prepis prvého, resp. posledného písmena. Pravidlo $(-, a, c) \rightarrow ab$ sa použije na prepis písmena *a* nachádzajúceho sa v pravidle v strede trojice v situácii keď je to prvé písmeno slova a za ním stojí písmeno *c* na reťazec *ab*.

a	-	a	b	c
-	ab		a	ab
a				
b	b		b	B
c	bb			

b	-	a	b	c
-	bb	ab	c	bb
a	bb	bb	b	bb
b	a	c	b	b
c	bc	bb	a	bb

c	-	a	b	c
-	bc	bb	b	
a	bc		b	
b	c		b	
c				

Ako ilustráciu uvedieme ako tento L systém regeneruje podreťazce *a* a *bb* reťazca ab^4c

$$a \Rightarrow ab \Rightarrow abb \Rightarrow aba \Rightarrow abbb \Rightarrow abba \Rightarrow abcb \Rightarrow abbbbc$$

V prvom kroku sme použili pravidlo tabuľky pre *a*, pred ktorým a za ktorým nestojí písmeno. V odpovedajúcom políčku je *ab*. V štvrtom kroku pri prepise *aba* sme použili pravidlá $(-, a, b) \rightarrow a$, $(a, b, a) \rightarrow bb$, $(b, a, -) \rightarrow b$. Získali sme reťazec *abbb*.

$$bb \Rightarrow ca \Rightarrow bbbc \Rightarrow cbbc \Rightarrow babc \Rightarrow abbbbc \Rightarrow abbbbc$$

Ilustrujeme tretí krok odvodenia, prepis reťazca *bbbc*. Tu sme použili pravidlá $(-, b, b) \rightarrow c$, $(b, b, b) \rightarrow b$, $(b, b, c) \rightarrow b$, $(b, c, -) \rightarrow c$. Získali sme reťazec *cbbc*.

Čo znamená zaoberať sa regeneráciou L systémov? Intuitívne požadujeme, aby bol reťazec vývojaschopný aj po poškodení a po čase, keď sa zregeneruje, sa vyvíjal tak, ako keby ku poškodeniu vôbec nedošlo. Pri snahe precízne formulovať túto úlohu pre L systémy sa vynárajú nasledujúce otázky. O regeneráciu ktorých reťazcov nám pôjde? Všetkých, alebo len niektorých? Napríklad dospelých, ktoré sa už z hľadiska vytvorenej štruktúry nemenia (aj keď môžu vykazovať vnútornú dynamiku)? O regeneráciu ako rozsiahlych poškodení nám pôjde? Budeme vyžadovať úplnú alebo len čiastočnú regeneráciu, teda regeneráciu z akéhokoľvek (drobného) zvyšku pôvodného reťazca, alebo len z niektorých reťazcov? Dovolíme, aby sa počas regenerácie vyskytli vo vývoji L systému aj iné symboly alebo len tie, ktoré využíva systém aj tak počas svojho „nenarušeného“ vývoja? Budeme hľadať tie reťazce, ktoré je schopný daný L systém regenerovať, alebo budeme cielene konštruovať / vytvárať L systémy s vysokou či predpísanou schopnosťou regenerácie, prípadne skúmať, čo umožňuje či spôsobuje schopnosť systému regenerovať? Ako ovplyvňuje veľkosť interakcie L systému jeho schopnosť regenerácie? Pozrime sa na príkladoch na niektoré odpovede na hore uvedené otázky.

Historicky sa ako prvá študovala možnosť regenerovať reťazec zložený z troch rôznych zhodne dlhých úsekov, teda reťazec v tvare $aa...abb...bcc...c = a^i b^i c^i$. Tento prípad je zvykom nazývať problém „francúzskej zástavy“ – *FZ*.

G. T. Herman vyžadoval od systému schopného regenerovať takéto reťazce, aby [11]:

1. Z reťazca zloženého výlučne z rovnakých symbolov, pred ktorými je špeciálny symbol *i* označujúci poškodenie reťazca, systém umožnil vytvoriť *FZ* rovnakej dĺžky akú má poškodený reťazec.
2. Reťazce zložené výlučne z rovnakých symbolov (t.j. bez špeciálneho symbolu *i*) sa aplikáciou pravidiel modelu nemenia.
3. Ak reťazec *FZ* rozdelíme na časti tak, že sa symbol susediaci s miestom delenia zmení na *i*, potom z takto zmeneného podreťazca (obsahujúceho aspoň dve bunky) opakovaným použitím pravidiel systému získame *FZ* rovnakého rozmeru ako bol podreťazec a s rovnakou orientáciou farieb.

G. T. Herman študoval symetrické[13] 2L systémy a navrhol pre problém regenerácie *FZ* symetrický, rozvíjajúci sa deterministický 2L systém, ktorý ho rieši (v hore uvedenom zmysle). Jeho systém sa dá zovšeobecniť na viacfarebné reťazce zostavené z rovnako

[13] 2L systém je symetrický ak spĺňa dve z nasledujúcich podmienok: (1) $(x, a, y) \rightarrow w$ je pravidlom systému súčasne s pravidlom $(y, a, x) \rightarrow w$. Hovoríme o vonkajšej symetrii. (2) ak $(x, a, y) \rightarrow w$ je pravidlom systému, potom sa *w* rovná svojmu zrkadlovému obrazu w^R. (3) $(x, a, z) \rightarrow w$ je pravidlom systému súčasne s pravidlom $(y, a, x) \rightarrow w^R$. Systém je polárne symetrický.

dlhých pruhov. Kým požiadavka na regeneráciu *FZ* znamenala obnovenie štruktúry *FZ* na kratšom reťazci, u reťazcov s bohatšou vnútornou štruktúrou je potrebné a často i nevyhnutné požadovať, aby sa regeneráciou obnovila ako štruktúra, tak aj pôvodná dĺžka reťazca.

Študujme zovšeobecnený problém, keď nás zaujíma možnosť regenerácie ľubovoľného reťazca. Od systému realizujúceho takúto regeneráciu reťazca w vyžadujeme, aby sa reťazec w aplikovaním pravidiel systému nemenil. V prípade, že w rozdelíme na podreťazce tak, že sa každé písmeno a susediace s miestom delenia zmení na $\bar{a}$, označujúce poškodenie, potom opakovaným použitím pravidiel systému na podreťazce u reťazca w, pre ktoré platí, že $u \neq u^R$, získame reťazec w alebo w^R. (^{R} reprezentuje zrkadlový obraz, t.j. slovo zapísané odzadu).

K. Ruohonen ukázal, že ku každému reťazcu w existuje symetrický, rozvíjajúci sa, deterministický 2L systém, ktorý reťazec w regeneruje [49].

Problému regenerácie sme sa venovali aj v Bratislave. Zaujímalo nás, či je možné dosiahnuť (úplnú) regeneráciu častí reťazca aj v prípade, že systém pracuje len so symbolmi pôvodného slova w, teda bez použitia nových symbolov signalizujúcich poškodenie. Podmienky, za ktorých takýto systém existuje, viedli k nutnosti použiť L systémy s rozsiahlejšou interakciou.

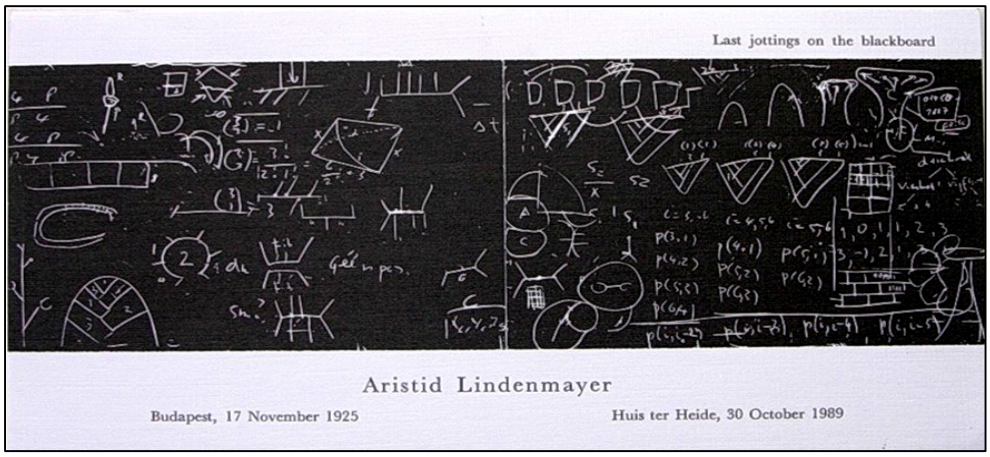
Problém regenerácie patrí medzi tie problémy, ktoré je jednoduché sformulovať a vysvetliť. Ako taký ho mohla začať skúmať aj študentka gymnázia A. Markoša v Bratislave, Zuzana Reichwalderová. S prácou o regenerácii reťazcov a^n, a^nb^n sa úspešne zúčastnila aj celoštátneho kola stredoškolskej odbornej činnosti v roku 1986. Na jej výsledky naviazalo štúdium regenerácie reťazca $a_1^{n_1}a_2^{n_2}\ldots a_r^{n_r}$ zovšeobecnenej *FZ*, ktoré sme prezentovali na konferencii v Salgótarjáne v máji 1986 [30]. Veľkosť interakcie získaných IL systémov, schopných úplne regenerovať slovo w ľubovoľnej štruktúry je závislá od dĺžky najdlhšieho podreťazca, ktorý sa v pôvodnom reťazci vyskytuje v dvoch rôznych polohách. IL systémy riešiace problém je možné nájsť v [25,26,27].

7 Výzvy súčasnosti

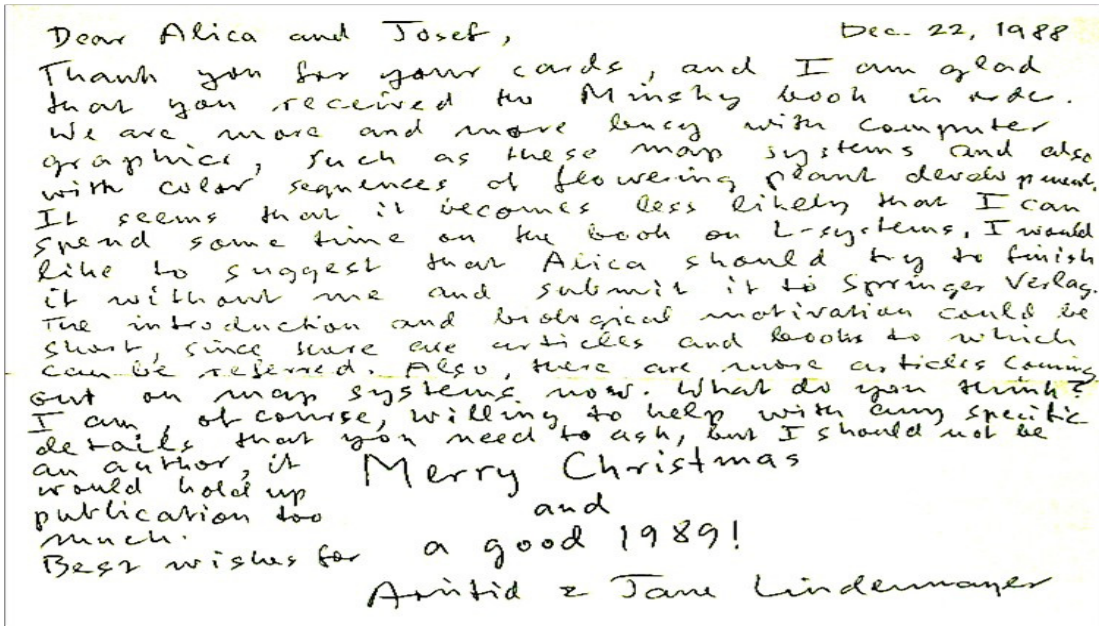
Každý, kto sleduje vývoj v oblasti Lindenmayerových systémov, musí byť v súčasnosti prekvapený nebývalým rozvojom oblasti a neuveriteľnou rozmanitosťou aplikácií L systémov, ktoré dokumentujú životaschopnosť, opodstatnenosť a dlhodobú aktuálnosť výskumu L systémov a aj gramatických štruktúr všeobecne. Takýto rozmach L systémov nemohol ich tvorca predvídať a prekvapuje určite aj všetkých jeho nasledovníkov. Ako ukážku z posledného obdobia môžeme uviesť určite netradičné aplikácie v muzikológii, napr. *hudobné L systémy* [2, 40,57], v architektúre [42,45] a počítačovej grafike pri animácii [15,16].

8 Záverom

Aristid Lindenmayer zomrel 30. októbra 1989 v holandskom Huis ter Heide. Na parte, ktoré rozosielala jeho priateľom, známym a kolegom jeho manželka Jane, americká výtvarníčka z New Yorku, s ktorou sa zoznámil počas svojho pôsobenia v USA, bola vyobrazená tabuľa v stave, v akom ju vo svojej pracovni A. Lindenmayer zanechal. Dokladá, že bol do svojich posledných dní plný nápadov a ochoty podeliť sa s nimi.



Vedel, že všetky sa asi nepodarí realizovať a dával to veľmi citlivo ľuďom, s ktorými mal určité plány, aj najavo. Žiaľ, človek si niektoré fakty uvedomí až *ex post* aj keď ich má vlastne pred sebou čierne na bielom:



Dear Alica and Josef, Dec. 22, 1988

Thank you for your cards, and I am glad that you received the Minsky book in time. We are more and more busy with computer graphics, such as these map systems and also with color sequences of flowering plant development. It seems that it becomes less likely that I can spend some time on the book on L-systems. I would like to suggest that Alica should try to finish it without me and submit it to Springer Verlag. The introduction and biological motivation could be short, since there are articles and books to which can be referred. Also, there are more articles coming out on map systems now. What do you think? I am, of course, willing to help with any specific details that you need to ask, but I should not be an author, it would hold up publication too much.

Merry Christmas and a good 1989!

Best wishes for

Aristid & Jane Lindenmayer

Literatúra

[1] Albert, M.H., Lawrence, J.: A proof of Ehrenfeucht´s conjecture. *Theoretical Computer Science* **41** (1985) 121-123.

[2] McCormack, J.: Grammar-based music composition *Complexity International* **3** (1996).

[3] Csontó, J., Palko M.: *Umelý život*. Košice: Elfa 2002

[4] Csuhaj-Varjú, E., Dassow, J., Kelemen, J., Paun, G.: *Grammar Systems – A Grammatical Approach to Distribution and Cooperation*. Yverdon: Gordon and Breach 1994.
[5] Csuhaj-Varjú, E., Kelemen, J., Kelemenová, A., Paun, G.: Eco-grammar systems – a grammatical framework for studying life-like interactions. *Artificial Life* **3** (1997) 1-28.
[6] Čulik II, K.: On the decidability of the sequence equivalence problem for D0L systems. *Theoretical Computer Science* **3** (1976) 75-84.
[7] Čulik II, K., Friš, I.: The decidability of the equivalence problem for D0L systems. *Information and Control* **35** (1977) 20-39.
[8] Čulik II, K., Karhumaki, J.: Systems of equations over a free monoid and Ehrenfeucht's conjecture. *Discrete Mathematics* **43** (1983) 139-153.
[9] Čulik II, K, Karhumaki, J.: A new proof of D0L sequence ekvivalence problem and its implications, In: *The Book of L* (G. Rozenberg, A. Salomaa, eds.). Berlin: Springer-Verlag 1986, pp. 63-74.
[10] Ehrenfeucht, A., Rozenberg, G.: Elementary homomorphism and a solution of the D0L sequence equivalence problem. *Theoretical Computer Science* **7** (1978) 169-183.
[11] Herman, G. T.: Models for cellular interactions in development without polarity of individual cells, I and II. *International Journal of System Science* **2** (1971) 271-289 and **3** (1972) 149-175.
[12] Herman, G. T., Rozenberg, G.: *Developmental Systems and Languages*. Amsterdam: North-Holland 1975.
[13] Honkala, J.: The D0L problem revisited. *Bulletin EATCS* **70** (2000) 142-147.
[14] Honkala, J.: The DF0L language equivalence problem. *Bulletin EATCS* **80** (2003) 143-152.
[15] Hornby, G.S., Pollack, J.B.: Creating high-level components with a generative representation of body brain evolution. *Artificial Life* **8** (2002) 223-146.
[16] Hornby, G.S., Pollack, J.B.: Envolving L systems to generate virtual creatures *Computers and Graphic* **25** (2001) 1041-1048.
[17] Chvál, J.: Lindenmayerove systémy ako efektívny nástroj modelovania morfogenézy rastlín. In: *Kognice a umělý život* (J. Kelemen, V. Kvasnička, J. Pospíchal, eds.) Opava: Slezská univerzita 2001, s. 59-69.
[18] Chvál, J.: Počítačom generované krajiny a umelý život. In: *Kognice a umělý život II,* (J. Kelemen, V. Kvasnička, eds.) Opava: Slezská univerzita 2002, s. 51-59.
[19] Kaandorp, J. A., Kubler, J. E.: *The algorithmic Beauty of Seaweeds, Sponges and Corals.* Heidelberg: Springer Verlag 2001.
[20] Karhumäki, J.: *An example of a PD2L system with the growth type 2½.* University of Turku. Department od mathematics, 1973.
[21] Karhumäki, J.: The decidability of the sequence equivalence problem for polynomially bounded D0L sequences. *Rev. Francaise Automat. Recherche Opérationnelle,* 1976.
[22] Kelemen, J.: *Myslenie, počítač…* Bratislava: Spektrum 1989.
[23] Kelemenová, A.: Levels in L-systems. *Mathematica Slovaca* **33** (1983) 87-97.
[24] Kelemenová, A.: Complexity of L systems – themes and études. In: *The Book of L* (G. Rozenberg and A. Salomaa, eds.). Berlin: Springer-Verlag 1986, pp. 179-191.

[25] Kelemenová, A.: Regeneration in IL systems. *Discrete Applied Mathematics* **22** (1989) 1-86.
[26] Kelemenová, A.: Size of context in regenerative IL systems. *Discrete Applied Mathematics* **26** (1990a) 117-119.
[27] Kelemenová, A.: Minimal context in regenerative IL systems. *Pure Mathematics and Applications (Serie A)* **1** (1990b) 329-336.
[28] Kelemenová, A., Kelemen, J.: An interview with A. Lindenmayer. *Bulletin of the European Association for Theoretical Computer Science* **23** (1984) 185-198.
[29] Kelemenová, A., Kelemen, J.: O biológii, matematike a výpočtoch – rozhovor s A. Lindenmayerom. *Pokroky matematiky, fyziky a astronómie* **31** (1986) 96-105.
[30] Kelemenová, A. Reichwalderová Z.: An IL model of regeneration. In: *Proc. Conf. Automata, Languages and Programming Systems* (F. Gécseg and I. Peák, eds.). University of Economics, Budapest, 1986, pp. 152-159.
[31] Kelemenová, A., Removčíková, M.: A0L- and CFG-size of languages. In: *Lindenmayer Systems* (G. Rozenberg and A. Salomaa, eds.). Berlin: Springer-Verlag 1992, pp. 177-182.
[32] Kelemenová, A., Vajgel, P.: Productions in stable 0L systems. In: *Developments in Formal Language Theory* (G. Rozenberg, A. Salomaa, eds.). Singapore: World Scientific 1994, pp. 102-110.
[33] Králová, M.: Constant ratio-functions of Lindenmayer systems. *Mathematica Slovaca* **35** (1985) 283-294.
[34] Králová, M.: A note on the ratio function in D0L systems. In: *Trends, Techniques, and Problems in Theoretical Computer Science* (A. Kelemenová, J. Kelemen, eds.). Berlin: Springer-Verlag 1987, pp. 127-137.
[35] Langton, C. G.: Artificial life. In: *Artificial Life, the Proceedings of an Interdisciplinary Workshop on the Synthesis and Simulation of Living Systems* (C. G. Langton, Ed.). Redwood City, CA: Addison-Wesley 1989, pp. 1-47.
[36] Lindenmayer, A.: Mathematical model of cellular interactions in the development. I. Filaments with one-sided inputs. II. Simple and branching filaments with two-sided inputs. *Journal of Theoretical Biology* **18** (1968) 280-299, 300-315.
[37] Lindenmayer, A.: An open problem. In: *Unusual Automata Theory (Research Report DAIMI PB 15)*. Department of Computer Science, University of Aarhus, Aarhus, 1973.
[38] Lindenmayer, A.: Models for multicelular development: Characterization, inference and complexity of L systems. In*: Trends, Techniques, and Problems in Theoretical Computer Science* (A. Kelemenová, J. Kelemen, eds.). Berlin: Springer-Verlag 1987, pp. 138-168
[39] Lindenmayer, A., Rozenberg, G. (eds.): *Automata, Languages, Development.* Amsterdam: North-Holland 1976.
[40] Manousakis, S.: Musical L-systems. Master Thesis. The Royal Conservatory. The Hague 2006.
[41] Meinhardt, H.: *The Algoritmic Beauty of Seashells*. New York: Springer-Verlag 2002
[42] Müller, P., Wonka, P., Haegler., S.: Procedural Modelling of Buildings. *ACM Transactions on Graphics* **25** (2006) 614 – 623.

[43] Nielsen N.: On the decidability of some equivalence problems for D0L systems. *Information and Control* **25** (1974) 166-193.
[44] Paz, A., Salomaa, A.: Integral sequential word functions and growth equivalence of Lindenmayer systems. *Information and Control* **23 (**1973) 313-343.
[45] Parish, Y.I.H., Müller, P.: Procedular Modelling of Cities. In: *Proc of ACM SIGGRAPH-01,* Computer Graphic Annual Conference Series, (E. Fiume, ed) New York: ACM Press 2001, pp. 301-308.
[46] Prusinkiewicz, P., Lindenmayer, A.: *The Algoritmic Beauty of Plants.* Berlin: Springer-Verlag 1990 (druhé vydanie 1996).
[47] Rozenberg, G., Salomaa, A. (eds.): *L systems*. Berlin: Springer-Verlag 1974.
[48] Rozenberg, G., Salomaa, A.: The Mathematical Theory of L Systems. New York: Academic Press 1980.
[49] Ruohonen, K.: *Some regenerative Lindenmayer models. Licentiate thesis 1974,* Turku.
[50] Ruohonen, K.: Equivalence problems for regular sets of word morphisms. In: *The Book of L* (G. Rozenberg, A. Salomaa eds.). Berlin: Springer-Verlag 1986, pp. 393-401.
[51] Salomaa, A.: L-systems – a device in biologically motivated automata theory. In: *Proc. 2nd International Symposium MFCS´73*. Bratislava: Mathematical Institute of the Slovak Academy of Science and UN Computing Research Centre 1973, pp. 147-151.
[52] Salomaa, A.: *Formal Languages*. New York: Academic Press 1973.
[53] Salomaa, A.: Events and languages In: *People & Ideas in Theoretical Computer Science* (C. S. Calude, Ed.). Berlin: Springer-Verlag 1999, pp. 253-273.
[54] Salomaa, A., Soittola, M.: *Automata-Theoretic Aspects of Formal Power Series*. Berlin: Springer-Verlag 1978.
[55] Valiant, L.G.: The equivalence problem for D0L systems and its decidability for binary alphabets. Centre for Computer studies. University of Leeds 1975, Technical report No 74.
[56] Wood, D.: *Grammar and L Forms – An Introduction*. Berlin: Springer-Verlag 1980
[57] Worth, P., Stepney, S.: Growing Music: Musical interpretations of L systems. In: *Applications on Evolutionary Computing* . LNCS Vol. 3449. Berlin: Springer Verlag 2005, pp. 545-550

Fundamentálne princípy kognitívnej biológie

Ladislav Kováč[1]

Abstrakt. Kognitívna biológia sleduje syntézu poznatkov viacerých vedeckých disciplín v rámci jednotného chápania života ako epistemického rozvinutia vesmíru (epistemický princíp). V súhlase s evolučnou epistemológiou považuje evolúciu za progresívny proces hromadenia poznania. Poznanie sa ukladá do konštrukcií organizmov. Štruktúrna komplexnosť tých konštrukcií, ktoré predstavujú zabudované znalosti, predstavuje epistemickú komplexnosť. Na rozdiel od evolučnej epistemológie, kognitívna biológia je založená na predpoklade, že molekulárna úroveň je fundamentálnou pre poznanie. Stojí na princípe minimálnej komplexnosti, ktorý stanovuje, že najúčinnejší spôsob štúdia akejkoľvek charakteristiky života je študovať ju na najjednoduchšej úrovni, na akej sa vyskytuje. Viaceré princípy kognitívnej biológie sú podobné extremálnym princípom fyziky a predstavujú tesné prepojenie medzi ontickou a epistemickou oblasťou. Z evidencie o vedomom prežívaní emócií ľuďmi vyplýva princíp minimalizácie utrpenia. Má význam pre také koncepty ako je onticita, fitnes, altruizmus, veda, budúcnosť ľudstva. Tento princíp, rovnako ako ostatné princípy kognitívnej biológie, je deskriptívny, nie normatívny.

Moderný molekulárny biológ by mohol parafrázovať básnika Popea tvrdením, že správnym objektom pre štúdium človeka je baktéria [1].

1 Úvod

Rozvoj molekulárnej biológie, ktorý sa začal v šesťdesiatych rokoch minulého storočia, je vzrušujúcim a zároveň aj hrozivým. Poskytuje nám stále prenikavejší pohľad do mechanizmov fungovania živej bunky. To umožňuje účinné zásahy do tohto fungovania s dôsledkami, ktoré už možno prekročili ľudskú schopnosť predvídania, hodnotenia a riadenia. Výsledky manipulácie s jadrom bunky budú mať asi dôsledky ešte závažnejšie než výsledky manipulácie s jadrom atómu.

Znepokojujúca je rastúca inštrumentalizácia molekulárnej biológie. Viac a viac sa veda mení na technovedu, ba až čisté inžinierstvo. Priority bádania sa stále viac reorientujú s ohľadom na techniku. V ohromne kompetitívnej oblasti výskumu, s komplikovanými metódami, inštrumentálne poznanie získava prioritu. Situácia v molekulárnej biológii má pritom paralely v iných oblastiach súčasnej vedy. Naše inštrumentálne poznanie, naša schopnosť manipulovať s vecami a udalosťami, izolovane a mimo ich kontextu, sú veľké, naše porozumenie komplexnosti, na ktorej to všetko spočíva, je malé a zaostáva. Navyše, skladajúc dohromady rozobrané veci a udalosti v nových umelých kombináciách, vytvárame systémy o nebývalej komplexnosti.

Diskrepancia medzi tým ako málo vieme v porovnaní s tým, ako veľa dokážeme robiť, je možno unikátnou v dejinách života na Zemi (nie však, ako bude naznačené neskôr, v histórii života vo vesmíre). Schopnosť vedieť a robiť predstavovala dosiaľ

---

[1] Prírodovedecká fakulta Univerzity Komenského, Bratislava. E-mail: kovacl@fns.uniba.sk
Po prvýkrát uverejnené v angličtine pod názvom Fundamental principles of cognitive biology v časopise *Evolution and cognition* **6** (2000) 51-69. Táto poznámka, tak isto ako ďalšie, čo nasledujú, nebola v pôvodnom texte.

jednotu, išlo o dve strany tej istej evolučnej mince. Dnes sa zdá byť naliehavou potrebou, aby sa povážlivý rozdiel medzi ľudskou schopnosťou poznať a konať, medzi kogníciou a správaním, zmenšil alebo odstránil a táto potreba volá po tom, aby sa prioritami vedeckého výskumu stalo úsilie o hlbšie porozumenie ľudskej kognícii a správaniu. Konrad Lorenz, ktorý neustále zdôrazňoval túto potrebu [2], by bol možno potešený pohľadom na súčasný rozmach kognitívnych vied.

Vo všeobecnosti, kognitívne vedy usilujú o odhalenie *mechanizmov* kognitívnych procesov. Nebezpečie inštrumentalizácie, ba aj zneužitia poznania, je veľké (Changeux, J. P., citované v [3]). V protiklade, možno trochu vágnom, ku kognitívnym vedám stojí epistemológia, ktorá sa usiluje o pochopenie *podstaty* poznania: čím ono je, ako sa vyvíjalo, aká je jeho vesmírna rola. Epistemológia bola tradične ťažiskovou disciplínou filozofie. Súčasné pokusy o jej „naturalizáciu“, hlavne pod označeniami genetická epistemológia [4] alebo evolučná epistemológia ([5,6], ako aj odkazy na literatúru tam uvedené) môžu byť chápané ako súčasť pokusov o vyňatie z filozofie ďalšej z jej konštitutívnych disciplín a jej premenu na samostatnú oblasť vedy.

Filozofia nevytvára poznanie, aj keď vždy bolo zámerom, aby tomu tak bolo; bola a je zdrojom emocionálne a sociálne účinných vier. Iba niektoré z nich môžu náhodne mapovať niektoré črty sveta do reality konštruovanej človekom. Toto konštatovanie nie je podceňovaním významu filozofie. Keďže ľudská túžba po zmysle a univerzálnom porozumení prýšti z hĺbok emocionálneho podložia, epistemológia bude naďalej súčasťou filozofie. Je prestaviteľné, že epistemológia ako súčasť vedy, by sa nemusela rozvinúť do podoby prírodných vied, ale skôr možno do podoby formálnych vied ako sú matematika, kybernetika, štatistická fyzika, teória hier.

Môže molekulárna biológia predstavovať časť, či dokonca základy, „naturalizovanej“ epistemológie? Namiesto toho, aby poskytovala len výkonné nástroje pre odhaľovanie mechanizmov fungovania mozgu, môže molekulárna biológia otvoriť novú cestu k porozumeniu *podstaty* kognitívnych javov, vrátane emócií ba aj vedomia? Akokoľvek by sa na prvý pohľad mohlo toto rozlíšenie javiť nepodstatným, v skutočnosti má zásadný význam. Ak by sa tak stalo, pojmy a metódy molekulárnej biológie, ktoré boli dosiaľ v kognitívnych vedách úspešné najmä na úrovni inštrumentálnej vedy, získali by iný význam. Prístup „zdola nahor“ k epistemologickým problémom, ktorého súčasťou je aj molekulárna biológia, bol nazvaný kognitívnou biológiou [7]. Vzhľadom k tomu, že využíva pojmy a argumenty termodynamiky, môže byť považovaný za nadstavbu bioenergetiky [7,8,9]. Niektoré pionierske idey boli formulované Goodwinom [10], ktorý aj označenie „kognitívna biológia“ prvý použil. Hlavný kredit však patrí Hansovi Kuhnovi. V jeho chápaní život od svojho vzniku, počínajúc sebakopírujúcimi nukleovými kyselinami, bol a je nepretržitým hromadením poznania [11,12].

Táto publikácia je krátkym náčrtom základných princípov, ktoré konštituujú kognitívnu biológiu. Ich počet nie je úplný. Hoci by sa zdalo, že niektoré majú normatívny charakter, bude dokazované, že všetky sú deskriptívne. Ich formálnejší a detailnejší výklad bude predmetom ďalších publikácií[2].

[2] Súčasná úroveň rozvinutia tohto prístupu bola načrtnutá v roku 2004 v publikácii *Komentovaný úvod do kognitívnej biológie*. V: Kelemen, J., Kvasnička, V., Pospíchal, J. (Eds.): Kognice a umělý život IV. Opava. Slezská univerzita, str. 233-258.

2 Rozvinutie koncepcií

2.1 Princíp logickej parity

Vzhľadom k záujmu biológov o rôzne stránky ľudskej kognície je prekvapujúce, ako málo pozornosti sa dosiaľ venovalo biológii logiky. Dokonca aj v štúdiách evolučných epistemológov sú takéto úvahy ojedinelé. Popper vo svojich ranných prácach trval na dôslednom rozlíšení logiky a psychológie pri analýze vedeckého skúmania [13]. Podľa Vollmera ([14] str. 180) povaha logiky je jeden z najťažších problémov a nie je zatiaľ rozriešený. Je možno predpokladať, že pre Lorenza bola logika výsledkom ľudskej evolučnej trajektórie, pritom ale odrážajúca správne relevantné črty sveta [15]. Na druhej strane v chápaní Piageta [4] logika nie je ľudskému indivíduu vrodená v tom zmysle, že by existovala v akejkoľvek fáze individuálnej ontogenézy. Logicko-matematické štruktúry sa vyvodzujú behom ontogenézy z operácií subjektu na prostredí; prestavujú čosi ako abstraktnú manipuláciu zrelého ľudského jedinca s objektmi jeho prostredia. Vari najdôkladnejšiu analýzu evolučnej povahy logiky poskytol Riedl [16,17]. Podľa neho evolúcia logiky je úzko zviazaná s evolúciou jazyka a logika je aj určená jazykom. Európska logika so svojou subjekt-predikátovou štruktúrou je podľa Riedla určená štruktúrou európskych jazykov a je iná ako napríklad čínska.

Proti názoru, že logika je „ľudsky špecifická“ či dokonca „kultúrne špecifická“, a že sú vo svete predstaviteľné početné vnútorne konzistentné, ale navzájom sa vylučujúce logiky, je možno argumentovať, že počítače, neživé stroje, sa riadia pravidlami formálnej logiky. K tomu by sa zas dalo trochu naivne namietnuť, že to tak nemusí byť, lebo počítače sú vynálezmi jedinej kultúry, tej istej, ktorá „vynašla“ formálnu logiku. Menej naivnou by sa javila námietka, že operácie počítačov musia byť izomorfné s operáciami ľudskej mysle, ktoré nie sú „kultúrne špecifické“, no sú „druhovo špecifické“, pretože počítače sú len „*exosomatickými nástrojmi*“ ľudského druhu – v analógii s mechanickými náčiniami a strojmi, ktoré Lotka označil ako „exosomatické nástroje“, a tento termín potom spopularizoval Georgescu-Roegen ([18] str. 307).

Bude neskôr ukázané evolučnou analýzou, že počítačová metafora ľudskej mysle je nepatričná a zavádzajúca[3], takže námietka s počítačmi je neopodstatnená. Je pravdepodobné, že v priebehu evolúcie motorické správanie organizmov, mechanická práca na prostredí, presúvanie objektov v priestore a čase, sa stávala stále menej zjavnou, bola postupne internalizovaná, viac a viac redukovaná len na naznačenie konania a transformovaná na vnútorné virtuálne manipulovanie s objektmi. Konečným vyústením tejto trvalo pribúdajúcej abstrakcie je ľudské myslenie – ako abstraktné motorické správanie. Vollmer ([14] str. 104), odkazujúc na Lorenza [19], považuje manipuláciu v imaginatívnom priestore („das Hantieren im Vorstellungsraum“) za počiatočnú fázu myslenia.

[3] Dôvody, prečo je počítačová metafora mozgu neadekvátna, boli analyzované v týchto publikáciách autora: (1) *Evolutionary primacy of emotions: the brain is a gland and psychoses are emotional derangement*. V: New York: Symposium on evolutionary theory and psychopathology 1999. (2) *Fedorovské inšpirácie*. V: Kognitívne vedy IV. Bratislava: ChF STU 2001, str. 39-46. (3) *Ľudské vedomie je produktom evolučnej eskalácie emocionálneho výberu*. V: Kelemen, J. (ed.): Kognice a umělý život III. Opava: Slezská univerzita 2003, str. 75-93. (4) *Life, chemistry and cognition*. EMBO Rep. 7 (2006) 562-566.

Objekt nemôže naraz byť i nebyť na tom istom mieste; dva objekty nemôžu naraz zaberať to isté miesto v priestore, atď. No toto je presne to, čo v abstraktných termínoch vyjadruje logika. Všetky pravidlá formálnej logiky nie sú ničím iným než abstraktným prekladom fyzikálnej nevyhnutnosti. To implikuje, že logické operácie našej mysle, aj našich počítačov, sú iba tautologickými *prekladmi*, ktoré slúžia na to, aby nám fakty sveta robili pochopiteľnejšími. V dôsledku týchto operácií sme schopní organizovať vnemy a pojmy a takto *konštruovať* realitu. Realita je naozaj druhovo špecifická, ba aj kultúrne špecifická, ba do istej miery aj individuálne špecifická, ale neexistuje nijaká druhovo špecifická ani kultúrne špecifická logika. Konštatovanie, že rozštiepenie, dualita a zároveň komplementarita sveta a mysle, je iluzórne, vyjadruje princíp logickej parity.

Logika ako fenomén rozštiepenia svet/myseľ je základňou, na ktorej sa vybudovala logika ako formálna veda so všetkými jej vetvami (ako logika relačná, modálna, temporálna, deontologická, atď.). Mala by sa vlastne nazývať *veda o logike*. Umožňuje nám vytvárať a triediť pojmy, ktoré uľahčujú naše porozumenie sveta (konceptuálny aspekt), ale aj našu manipuláciu so svetom (inštrumentálny aspekt). Kognitívna biológia požaduje jasné definície pojmov zvolených na základe kritéria ich maximálnej konceptuálnej a inštrumentálnej užitočnosti, s jednoznačným ustanovením, že všetky pojmy sa vzťahujú na realitu a žiaden z nich nie priamo na svet.

Bolo by asi trochu trúfalé označiť princíp logickej parity za princíp všeobecnej komplementarity. Princíp komplementarity zaviedol Niels Bohr, aby vysvetlil dualitu vlna/častica v kvantovej mechanike. My ľudia žijeme vo svete stredných rozmerov, makrosvete. (Toto je termín, štandardne používaný vo fyzike. Odpovedá termínu „mezokozmos“, ktorý navrhol Vollmer [20]). Naša myseľ vníma fenomény makrosveta a vhodne ich konceptualizuje, ale nie je skonštruovaná tak, aby vnímala fenomény mikrosveta, ktorý je popisovaný matematickým formalizmom kvantovej fyziky. Aby sme tento formalizmus previedli do podoby „šitej na mieru človeka“, musíme sa uchýliť k perceptom a konceptom makrosveta. To je dôvod, prečo si predstavujeme elektrón ako časticu pri vysvetľovaní výsledkov z jedného pokusného usporiadania a ako vlnu pri interpretácii dát z iného usporiadania.

Podľa Bohra [21] musíme použiť ten istý princíp komplementarity, aby sme vysvetlili dualitu sveta a mysle. Sledujúc rovnakú argumentáciu, mohli by sme svet, ktorého sa táto dualita týka, označiť ako „psychosvet“. Zdá sa, že rovnaký princíp treba použiť aj pre vysvetľovanie „megasveta“, sveta galaxií, ktorý je predmetom kozmologického skúmania. Komplementarita psychosveta asi vyplýva najmä z jeho „mnohorozmernej“ komplexnosti. To je dôvod, prečo by mohol byť užitočný princíp komplementarity pri výklade ešte iného sveta, „sociosveta“, výtvoru kultúrnej evolúcie. Aj ten sa asi svojou komplexnosťou vymyká jednoznačnému ľudskému popisu a možno potrebuje viacero komplementárnych vysvetlení, každé vnútorne konzistentné, a pritom sa vylučujúce s inými. Pod všetkými týmito komplementaritami je vlastne dualita svet/myseľ.

Princíp logickej parity je dôležitý pre pojem racionality. Bude ukázané neskôr, že je užitočné rozlišovať konceptuálnu a inštrumentálnu racionalitu. Každá z nich nepredstavuje stav, ale proces. U oboch ide o to spojiť dva body trajektórie. Ideálna racionalita odpovedá najkratšej trajektórii, ktorá vedie k cieľu, v najjednoduchšom prípade priamke. Úplnému chýbaniu racionality odpovedá náhodné blúdenie. „Reálna“ racionalita, ktorá zahrňuje „ohraničenú“ racionalitu ľudí [22], leží medzi obidvoma extrémami a je

určená stupňom poznania, ktoré je k dispozícii. Z toho je zrejmé, že pojem racionality je paritný s pojmom „extremálnych princípov“ vo fyzike, na ktorých možno spočívajú všetky fundamentálne zákony prírody [23]. Táto parita má možno hlboký zmysel, ktorý ale nebol dosiaľ dostatočne preskúmaný.

Ak fyzikálna nevyhnutnosť môže byť vždy preložená do logickej nevyhnutnosti, je možný aj opačný preklad (až na to, že logika sa zaoberá všetkými možnými svetmi, nielen tými, ktoré vytvorili kontingencie evolúcie nášho vesmíru). Čo je logické, je vždy možné, buď virtuálne, v princípe, alebo ako fakt prírody. Darwinovská interpretácia evolúcie sa všeobecne považovala za teóriu, ktorá je užitočná pri objasňovaní našich pozorovaní živej prírody. Nie je to teória. Je to logická nevyhnutnosť, vesmírny fakt v akomkoľvek vesmíre, v ktorom sa vyskytujú sebareplikujúce entity. V každom prostredí, ktoré obsahuje obmedzené množstvo zdrojov, automaticky musí prebiehať exponenciálne množenie, súťaž a výber medzi sebareplikujúcimi entitami. Biologická evolúcia je takto len dielčim prípadom faktu evolúcie [24]. Replikátorová rovnica [25] je jednou z najfundamentálnejších rovníc vesmíru.

Ako však dokázať, že náš vesmír, so svojou logikou sebareplikujúcich entít, je skutočný a nie iba virtuálny? To bude predmetom posledného z princípov, ktoré analyzuje táto štúdia.

2.2 Princíp dvojitého epistemického uzavretia

Je všeobecnou charakteristikou akéhokoľvek živého systému, že zaznamenáva relevantné črty svojho prostredia a na ne adekvátne reaguje tak, aby si udržal svoju permanenciu v danom prostredí, *onticitu.* Vlastne význam konkrétnych vlastností prostredia je určený ich hodnotou, pozitívnou alebo negatívnou, akú majú pre udržovanie onticity konkrétneho živého systému; tie, ktorých hodnota je neutrálna, nie sú zaznamenávané. Ľudský druh sa po dlhú dobu nelíšil v tejto charakteristike od iných biologických druhov, až kým ohromná redundancia ľudského mozgu neumožnila kultúrnu evolúciu. Ako kultúrna evolúcia postupovala, mýty, druhovo-špecifický znak človeka, zaisťujúci onticitu potlačovaním kognitívneho zmätku a podporovaním skupinovej kohézie, sa premieňali na filozofiu a neskôr na experimentálnu vedu. Len čo sa raz ustanovili, filozofia a veda svojimi autonómnymi dynamikami, v podstate nezávislými na ľudských úmysloch, urobili z neutrálnych vlastností prostredia, ktoré nemali biologickú relevanciu, objekty ľudského zaznamenávania a hodnotenia a zdroje nových foriem správania. Sebauvedomovanie, samo produkt redundancie mozgu, dostalo novú, ďalšiu funkciu: reflexiu sveta a reflexiu o svete.

Pri veľmi hrubom a silno zjednodušujúcom pohľade sa zdá, že pre väčšinu ranných filozofov v rodisku filozofie, antickom Grécku, ľudská reflexia sveta nemala obmedzenia, akokoľvek obmedzenými a povrchnými sú bezprostredné vnímanie a kognícia. *Logos* prírody bol považovaný za izomorfný s *Logom* ľudskej mysle, a teda úplne prístupný ľudskému pochopeniu. Postupne sa k tomuto názoru objavovali alternatívy, váhania a pochybnosti, ktoré vyvrcholili v Kantovej transcendentálnej filozofii. Kantov názor na hranice, uložené ľudskému poznávaniu, biologicky reinterpretoval Konrad Lorenz [26]. Túto reinterpretáciu, ktorú anticipovali viacerí Lorenzovi predchodcovia (uvedení v [27]), možno považovať za *základný kameň* evolučnej epistemológie. Predkantovský epistemologický skepticizmus, napr. v podobe štyroch „idolov“ Francisa Bacona, by bolo možno reinterpretovať a „naturalizovať“ podobným spôsobom. Obmedzenia ľudskej

mysle dôkladne analyzovali všetci evoluční epistemológovia. Riedl tak urobil vo viacerých knihách a stručne v jednej časopiseckej publikácii [28].

Hranice ľudskej mysle, jej možnosti a obmedzenia, uložené kontingenciami biologickej evolúcie ľudského druhu, sa zdajú byť neprekročiteľné. V dôsledku toho je naša realita, model sveta, druhovo-špecifickou. (Z čoho plynie, že iné biologické druhy majú svoju vlastnú druhovo-špecifickú realitu.) Uzatvárajú nás do sveta stredných rozmerov a malej komplexnosti. Svet za týmito hranicami je od nášho sveta oddelený bariérami, ktoré možno nazvať *Kantovými bariérami*. Keď sa pokúšame prekročiť tieto bariéry, je princíp komplementarity jediným prostriedkom ako interpretovať tie javy, s ktorými sa tam stretávame. Druhovo-špecifické vymedzenie kognitívnych kapacít možno nazvať prvým epistemickým uzavretím.

Je ale aj iné uzavretie, všeobecnejšie a fundamentálnejšie. Anticipovali ho mnohí filozofi, počínajúc menšinou filozofických „disidentov" v gréckej antike (Protagoras, Timon, Zenón), cez Humea po Kanta. Rigoróznu formuláciu mu dal v r. 1931 Kurt Gödel svojou teorémou o neúplnosti. Gödelova teoréma sa týka formálnych systémov matematiky a obmedzení, ktorým podliehajú (jednoduchý výklad je v [29,30,31])[4]. Akúkoľvek množinu konzistentných pravidiel použijeme pre manipuláciu s matematickými symbolmi v systéme axióm, vždy existuje tvrdenie, vyjadrené v jazyku týchto symbolov, ktoré nemožno s použitím daných axióm a pravidiel ani dokázať ani vyvrátiť. Navyše, v uzavretom systéme konečného počtu axióm a pravidiel, nie je ani možnosť povedať, či východzie predpoklady sú logicky konzistentné alebo nie. Ak sa pokúsime vyriešiť problém pridaním nového pravidla alebo novej axiómy, vytvoríme nové nerozhodnuteľné tvrdenie. Aby sme plne porozumeli systému matematiky, musíme systém prekročiť.

Ako poznamenal Hofstadter, aplikovanie Gödelovej teorémy na iné oblasti môže mať istú cenu v prípade, ak aplikáciu chápeme metaforicky a neberieme ju doslovne ([30] str. 696). Matematici bývajú znechutení a podráždení, keď sa nematematici odvolávajú na Gödelovu teorému, aby zdôvodnili svoj skepticizmus, vzťahujúci sa na mimomatematické oblasti. Ako však ukázal Chaitin [32], Gödelovu teorému netreba chápať ako izolovaný paradox, ale ako prirodzený dôsledok obmedzení, ktoré na kognitívnu kapacitu ukladá teória informácie. V teórii algoritmickej komplexnosti [33,34] komplexnosť radu znakov sa rovná počtu bitov minimálneho programu pre zhotovenie tohto radu. Systém môže byť predstavovaný veľmi dlhým radom znakov, ale jeho komplexnosť je malá, ak program, ktorým možno tento rad znakov zhotoviť, sám je tvorený len malým počtom znakov. Informácia radu je prítomná v „stlačenej" podobe v minimálnom programe. Ak je minimálny program rovnakej dĺžky ako je rad, jeho informácia nemôže byť komprimovaná, pretože rad znakov je randomný[5]. A keďže minimálny program nemôže byť komprimovaný, je vždy randomný. V nejakom systéme axióm a pravidiel inferencie je možno vždy dokázať, že rad je nerandomný – tým, že sa nájde program na jeho zhotovenie, ktorý má menej znakov – ale neexistuje spôsob ako dokázať, že je randomný,

[4] Na Slovensku sa Gödelovým vetám a matematickým paradoxom venuje Pavol Zlatoš. Jeho kniha má názov *Ani matematika si nemôže byť istá sama sebou*. Bratislava: Iris 1995.

[5] Anglické slovo „random" sa do slovenčiny niekedy prekladá ako „náhodný". Kvôli jednoznačnosti zdá sa byť výhodnejším slovo neprekladať a do slovenčiny zaviesť novotvar „randomný". Slovo „informácia" je vo výklade Chaitinových názorov ponechané, ale samotné chápanie informácie je predmetom kritickej analýzy autorom: *Information in biology: A time for rethinking the fundamentals*. Plant Signaling & Behavior, v tlači

a že ho už viac stlačiť nemožno. Počítačový program, ktorý by bol spustený, aby toto dokázal, by sa nikdy nezastavil. Vo formálnom systéme o istej komplexnosti nemožno dokázať pre nijaký rad znakov, že je randomný (teda o minimálnej algoritmickej komplexnosti), pokiaľ komplexnosť radu je menšia ako je samotný systém.

To rovno ukazuje, že Gödelova teoréma má vzťah k Turingovmu problému zastavenia ([30] str. 425), Tarskiho teórie pravdy [35] a k Popperovmu princípu falzifikácie [13]. Vlastne všetky tieto formulácie znamenajú to isté konštatovanie o systéme o istej komplexnosti (čo musí platiť nielen pre algoritmickú komplexnosť, ale aj pre iné chápania komplexnosti). Je nemožné – a nemožnosť je založená na logike, a teda s prekladom, na vlastnostiach materiálneho sveta – dosiahnuť úplné poznanie systému o špecifikovanej komplexnosti prostriedkami, dostupnými iba v rámci systému; aby sme to dosiahli, potrebujeme širší systém o väčšej komplexnosti, *metasystém*. Veľkosť komplexnosti formálneho systému určuje veľkosť a hranice jeho *epistemickej kapacity*.

Penrosovu snahu ukázať, že vedomá činnosť ľudského mozgu presahuje možnosti počítača [31,36] možno chápať v tomto kontexte. Ľudská myseľ je možno komplexnejšia ako je komplexnosť formálnych systémov matematiky. Ak ale tieto systémy a logika, ktorá je v nich, sú preložiteľné do vlastností fyzikálneho sveta, prekročenie kapacity týchto formálnych systémov by podľa Penrosa implikovalo, že nejaké nové, dosiaľ neznáme fyzikálne princípy, určujú nad-počítačovú schopnosť ľudského mozgu. Už Kuhlenbeck [37] vytušil hlboký vzťah medzi Gödelovou teorémou o neúplnosti a medzi čímsi, čo on nazval „paradoxom mozgu“. Poukázal na to, že náš svet vedomia je fenoménom mozgu, ale samotný mozog je tiež fenoménom mozgu. Teda je tam uzavretie do kruhu. Aby sa paradox prekonal, bolo by treba iný hypotetický mozog, ktorý by bol umiestnený nie v doméne vedomia, ale v mimomentálnom verejnom časopriestorovom systéme. Gerlach [38], ktorý upozornil na túto príbuznosť medzi Gödelom a Kuhlenbeckom, navrhol premenovať paradox mozgu na „Kuhlenbeckov paradox“.

Aj Hofstadter analyzoval tento problém. Vyslovil presvedčenie, že nie je nijaká fundamentálna príčina, ktorá by bránila úplnému porozumeniu ľudskému mozgu. Mohol by sa javiť „úplne jasným inteligentnejším bytostiam“ ([30] str. 707). Nanešťastie, zatiaľ biologická evolúcia na Zemi takéto bytosti nevytvorila.

Princíp druhého epistemického uzavretia je fundamentálne iný ako princíp prvého uzavretia: Konštatovanie, že nejaký systém nemôže byť úplne porozumený prostriedkami systému samotného, platí pre každý formálny systém, akákoľvek by bola jeho komplexnosť. Vzťahuje sa aj na každý živý systém, nezávisle od toho, aké evolučné kontingencie tvarovali jeho kognitívne kapacity, a teda aj jeho druhovo-špecifickú realitu. Je to princíp logiky, nie evolúcie. Oba princípy epistemického uzavretia, prvého i druhého, sú vlastne už implicitné v Kantovej epistemológii.

Namiesto privolávania nejakého zázračného „metamozgu“ potrebného na to, aby sme dosiahli čo najdôkladnejšie porozumenie podstate ľudskej kognície, môžeme použiť prístup opačný: získať najprv úplné poznatky o kognícii živých bytostí, ktoré sú jednoduchšie ako človek. Keďže komplexnosť jednoduchších živých bytostí je určite menšia ako komplexnosť človeka – a toto sa zrejme vzťahuje aj na epistemické „formálne systémy“ týchto jednoduchších organizmov – ľudský kognitívny systém možno chápať ako „metasystém“ vo vzťahu ku kognitívnym systémom jednoduchších druhov. Potom by extrapolácia týchto poznatkov až na ľudskú kogníciu mohla byť spôsobom akým obísť „*Kuhlenbeckovu bariéru*“. Z týchto úvah vyplýva ďalší princíp.

2.3 Princíp minimálnej komplexnosti

Problém, ako definovať komplexnosť, nie je rozriešený. To isté slovo sa používa na popis rozdielnych charakteristík rozdielnych systémov. V snahe vyhnúť sa zmätku je v tejto publikácii slovo komplexnosť používané vždy s prídavným menom tam, kde sú dostupné jednoznačné, pokiaľ možno matematické definície, napr. algoritmická, termodynamická, epistemická komplexnosť. Tam, kde je slovo použité bez prídavného mena, bude znamenať intuitívny, aj keď dosť vágny pojem na označenie niečoho, čo nie je jednoduché, ale zložené, organizované, komplikované[6].

Iba málo z toho, čo vieme o ľudskej dedičnosti, bolo získané štúdiom ľudí. Väčšia časť je výsledkom extrapolácií zo štúdií jednoduchších organizmov. Od hrachu zakladajúceho otca genetiky Mendela *smerom nadol* k ovocnej muške Morgana, ešte nižšie ku kvasinkám Ephrussiho a Lindegrena a k baktériám Ledeberga a Monoda, až celkom na dno k fágu Benzera, Delbrücka a Lvoffa, to bola víťazná cesta genetiky a molekulárnej biológie *smerom nahor*.

Max Delbrück, ktorý bol úspešným fyzikom predtým, než sa začal venovať biológii, zvlášť poznačil túto trajektóriu. Jeho ambíciou bolo popísať biologické systémy s rovnakou precíznosťou aká sa používa vo fyzike. Bolo mu zrejmým, že k tomu, aby to dosiahol, musí si k štúdiu zvoliť biologické systémy čo najjednoduchšie. Preto skúmal fototaxiu jednoduchej huby (*Phycomyces*) ako model organizmov, ktoré reagujú na vonkajšie podnety. Historickým sa stalo jeho rozhodnutie v r. 1937 začať študovať bakteriofága: možno ho považovať za začiatok molekulárnej biológie [39]. Bakteriofága možno považovať za najjednoduchšiu živú „vec“. Výskum bakteriofága umožnil vypracovanie koncepcie génu ako jednotky dedičnosti, vhľad do vnútornej štruktúry génu cez intragénové mutácie, vyjasnenie povahy genetickej rekombinácie. To všetko bolo uskutočnené pred tým, ako sa nukleovým kyselinám pripísala rola materiálnych nosičov génov a včaššie ako biochémia zaujala dominujúce postavenie v molekulárnej biológii.

Delbrückov úspech v genetike ako dôsledok skúmania najjednoduchšieho systému čo ešte javí dedičnosť, inšpiroval jeho spolupracovníkov a žiakov, aby si zvolili podobný prístup v inej oblasti vedy, v neurobiológii. Študovaná bola *Aplysia*, organizmus len s tisíckami neurónov (na rozdiel od ľudských miliárd) a neskoršie *Caenorhabditis elegans* len so stovkami neurónov. V r. 1998 bola publikovaná kompletná sekvencia všetkých génov v *C. elegans* [40] a hneď prvé porovnávacie analýzy génov kódujúcich neuronálne funkcie objavili prekvapujúcu homológiu s mnohými neuronálnymi génmi cicavcov a s génmi ľudských neurologických chorôb [41].

Táto skúsenosť, dokázaná aj v iných odvetviach vedy, opodstatňuje formovanie princípu, ktorý má heuristický i konceptuálny význam: Najúčinnejší spôsob pre štúdium konkrétneho biologického fenoménu je študovať ho na najjednoduchšom organizme, u ktorého sa tento fenomén vyskytuje – tam je experimentálne najprístupnejší a vzhľadom k jeho evolučnej jednoduchosti aj teoreticky najľahšie uchopiteľný.

To je princíp minimálnej komplexnosti. Aby sa ocenila Delbrückova zásluha, možno ho nazývať aj Delbrückovým princípom. Princíp by sa ľahko použil, ba priam automaticky, keby sme vedeli ľahko určiť evolučnú úroveň, na ktorej sa konkrétny fenomén objavil. No nájsť najnižšiu úroveň nie je ľahkým problémom. Problém súvisí s otázkou homológie a analógie. Dva fenomény môžu vyzerať rovnako, no pritom mať

[6] Rozdiel medzi „komplexnosťou“ a „komplikovanosťou“ je zásadný. Diskutovaný bude v inej publikácii.

rozdielny evolučný pôvod. Aj keby boli priam totožné svojou formou i funkciou, ich konceptuálne zjednotenie ostane iba metaforou. Metafora sa môže stať zavádzajúcou a viesť k naivným interpretáciám. Často je tomu tak ak fenoména nižšej úrovne sa vysvetľujú pomocou fenoménov vyššej úrovne, čo je štandardný prípad antropocentrizmu.

Podstata problému sa stáva jasnou, keď hľadáme homológie na úrovni vyššej ako je molekulárna. Do akej miery je sociálne správanie cicavcov homologické so sociálnym správaním hmyzu alebo baktérií? Je ľudské vedomie homologické s vedomím iných živočíchov? Ak ľudskému vedomiu pripíšeme 100%, má šimpanz 80%, myš 2%, ovocná muška 0,01% a kvasinka 0,000001% vedomia? Otázka siaha až na samotnú podstatu kognitívnych vied: je kognícia charakteristikou všetkých živých bytostí, rovná sa život kognícii, alebo, naopak, je kognícia výlučne charakteristikou človeka? Je podivuhodné, že oba extrémne pohľady existujú v súčasných kognitívnych vedách a majú svojich zástancov a bojovníkov. Aj v samotnej evolučnej epistemológii sa názory rozchádzajú ([42], [43] str. 294). Všetky aspekty tejto principiálnej otázky sú príliš rozsiahle a budú predmetom inej publikácie[7].

Podstatou problému je zrejme jav, ktorý je známy pod rôznymi menami: emergencia, fulgurácia, diskontinuita, kvalitatívny prechod. Ak v prípade vedomia nepripustíme, že existuje prah pre evolučný pôvod vedomia, skončíme v panpsychizme, pripisujúcom vedomie dokonca aj elementárnym časticiam hmoty. Niektoré kvalitatívne prechody, ako fázové prechody v termodynamike alebo perkolácie v štatistickej fyzike, majú exaktný matematický podklad. Na druhej strane veda popisuje niektoré kvalitatívne zmeny ako iba zdanlivé. Je to prípad vizuálneho vnímania, kde kvalitatívne diskontinuity farieb odpovedajú plynulým zmenám vlnovej dĺžky elektromagnetického žiarenia. Práve v tejto oblasti existuje hlavná výzva pre epistemológiu: niektoré „fázové prechody“ sa vyskytujú vo svete a sú prekladané do podoby „fázových prechodov“ v druhovo-špecifickej realite; iné sa týkajú javov ktoré nemajú svoj odraz v realite; a ešte iné – možno väčšina má takúto podobu – sa vyskytujú iba v realite a vnášajú do nej kvalitatívne odlišné fenomény, ktoré nemajú paralely vo svete (a toto sa stáva ešte komplikovanejším vzhľadom k povahe pojmov: všetky pojmy sú navzájom oddelené, vnucujú diskontinuity a rozčleňujú, niekedy celkom arbitrérne, realitu). Rozlíšenie medzi týmito tromi rozdielnymi kategóriami kvalitatívnych prechodov je záležitosťou empirického výskumu, a je teda vždy len provizórne. To je príčina, prečo vybrať výskumný objekt s minimálnou komplexnosťou je vždy záležitosťou neistou a jeho adekvátnosť sa dokáže až neskôr podľa výsledkov výskumu.

V každom prípade princíp predpokladá, že existujú rozdielne úrovne komplexnosti živého sveta, a že v priebehu biologickej evolúcie komplexnosť živých foriem kontinuálne rástla.

2.4 Epistemický princíp

Už od čias Darwina neprestali v biológii diskusie o tom, či biologická evolúcia je progresívna a má smer. Ak je v biologickej evolúcii pokrok, potom možno hovoriť o jednoduchších a zložitejších, nižších a vyšších organizmoch, možno sa pokúšať o nájdenie a definovanie evolučných tendencií, ba aj formulovať prípadné zákony

[7] Problém je analyzovaný v práci *Ľudské vedomie je produktom evolučnej eskalácie emocionálneho výberu*. V: Kelemen, J. (ed.): Kognice a umělý život III. Opava: Slezská univerzita 2003, str. 75-93.

evolúcie. Ak pokrok nie je, takéto označenia nemajú zmysel a musia byť odmietnuté ako antropomorfizmy.

Podstatou argumentácie tých, ktorí neuznávajú nijaký pokrok v evolúcii, je konštatovanie, že v špecifickom prostredí organizmy – alebo, ak použijeme „pohľad génovými očami“, ich gény – súťažia s inými organizmami o jednoduchú utilitu: darwinovskú fitnes. Fitnes sa dosahuje rozmanitými spôsobmi. V tom istom prostredí rôzne organizmy môžu používať rôzne stratégie a mnohé stratégie môžu byť rovnako účinné. Prostredím pre individuálny organizmus nie sú len prírodné podmienky, ale aj iné organizmy. Takže fitnes individuálneho organizmu je vysoko dynamickou funkciou, v ktorej premennými sú aj fitnes všetkých ostatných organizmov. Človek sa zdá byť organizmom s vysokou fitnes, so súčasnou populáciou vyše 6 miliárd indivíduí. Akokoľvek môže toto číslo vypadať impozantné, je úplne nepatrné v porovnaní s počtom, ale aj bunkovou hmotou, baktérií. Niektoré bakteriálne druhy existujú na Zemi možno nie príliš zmenené už niekoľko miliárd rokov a zdá sa byť pravdepodobným, že budú existovať aj potom, keď druh *Homo sapiens* už tu viac existovať nebude. Ak teda baktérie prežili tak dlho a majú pred sebou skvelú budúcnosť, niet dôvodu vidieť v evolúcii akýkoľvek pokrok a označovať ľudí za vyššie a baktérie za nižšie organizmy.

Kontroverzia okolo pokroku v evolúcii má svoj pôvod v nedorozumení, vyplývajúcom z dvojznačnej konotácie slova „pokrok“. V európskej kultúre, prinajmenšom od čias osvietenstva, pokrok je považovaný za niečo čo si treba priať, za čosi hodnotné, „dobré“. Bežne sa pokrok považuje za nevyhnutný znak kultúrnej evolúcie: čokoľvek, čo je pokročilejšie (pokrokovejšie?), má väčšiu šancu sa presadiť a pretrvať, má teda, dalo by sa povedať, väčšiu darwinovskú fitnes v porovnaní s niečím, čo je konzervatívne a retardujúce.

Keď ale opustíme normatívnu konotáciu, slovo „pokrok“ ostáva ako neutrálne, bezhodnotové vyjadrenie evolučnej tendencie. Tendencia je nespochybniteľná: je to tendencia k objavovaniu sa stále komplexnejších organizmov. Samotný fakt, že baktérie, s ich jednoduchou organizáciou a jednoduchým správaním, sú evolučne staršie ako človek, tvor s mozgom o najväčšej štruktúrnej komplexnosti aká sa dosiaľ v evolúcii vyskytla, je toho dôkazom. Tento fakt nič nemení na možnosti, že fitnes baktérií je väčšia ako fitnes človeka. Je aj možné, že niektoré baktérie sú evolučne mladšie ako je človek a pritom zároveň majú oveľa menšiu komplexnosť. No takáto možnosť nie je argumentom proti univerzálnej evolučnej tendencii, ktorú možno vyjadriť metaforou komplikovaného *bludiska*: život nepretržite, na všetkých úrovniach, miliónmi druhov, „testuje“ možnosti ako postúpiť dopredu. Ohromná väčšina druhov hynie alebo končí v slepých uličkách, v ktorých nie je viac možnosť ísť dopredu. Ísť dopredu – k akému cieľu?

Pokrok v evolúcii má svoju termodynamickú príčinu. Dynamika sveta je ireverzibilná, určená druhou vetou termodynamiky. Bez jej platnosti by svet mal newtonovský charakter: darwinovské variácie by sa podobali nekonečným hudobným variáciám na tú istú tému. Prirodzený výber by hral reverzibilne s týmito variáciami.

Ako pred časom zdôraznili Fontana a spol. ([44] str. 210), znovu oživujúc stále sa vracajúcu otázku súčasnej biológie, „výber nemá generatívnu silu; jedine vylučuje to, čo nie je ‚fit‘, takže sa týka *kinetického* aspektu evolučného procesu. Hlavný problém evolúcie je problém *konštruovania*: porozumieť tomu, ako vznikajú organizované formy, ktoré sú vystavené prirodzenému výberu a tomu, ako mutácie vedú k organizačným, teda fenotypickým, inováciám.“ Odpoveď v zásade už bola poskytnutá pred tromi

desaťročiami termodynamikou nerovnovážnych systémov: systémy ďaleko od rovnováhy sa vyvíjajú spontánne do usporiadaných stavov, nazvaných *disipatívne* štruktúry, v dôsledku fluktuácií [45]. Samozrejme, nie je to úplná odpoveď: *konštrukcie*, aké sa vyskytujú v živých systémoch, sú najmä *konzervatívnymi* štruktúrami, nie disipatívnymi a úlohou ostáva vyjasniť, ako tieto konštrukcie, stále zložitejšie, vznikajú a ako sú udržované. Práve tu vystupuje princíp minimálnej komplexnosti do popredia.

Často sa tvrdilo, že vznik života[8] musel byť veľmi nepravdepodobnou udalosťou, ku ktorej možno došlo iba raz v dejinách vesmíru. Podpora tomuto názoru sa celkom chybne hľadala v druhej vete termodynamiky. Prehliadalo sa, že druhá veta termodynamiky sa uplatňuje vo svete, ktorý je ovládaný fundamentálnymi silami. V poli síl elementárne častice, atómy, molekuly sa spájajú dohromady, čím vytvárajú vyššie jednotky, takže disipácia potenciálnej energie má tvorivú rolu: generuje štruktúry. Dnes sa zdá byť temer isté, že v ktorejkoľvek časti vesmíru, kde sú termodynamické podmienky teploty a tlaku podobné ako boli pred štyrmi miliardami rokov na Zemi, život musí vzniknúť ako fyzikálna nevyhnutnosť.

Keď sa objavili nukleové kyseliny schopné sebakopírovania, vo vesmíre sa vynorila nová kvalita: *molekulárne rozpoznávanie*. Medzi dvomi atómami vodíka a kyslíka, ktoré sa za špecifických podmienok spojujú a vytvárajú vodu, nie je nijaké rozpoznávanie. Toto ich spojovanie, asociácia, je fyzikálnou nutnosťou, je *nomickým* procesom. Ale asociácia molekúl nukleotidtrifosfátov s jednovláknovým templátom sebareplikujúcej nukleovej kyseliny, ktorá vedie k Watson-Crickovmu párovaniu k sebe patriacich báz, je iným typom procesu. Je to nutnosť *ex post*, pseudo-nomický proces, za ktorým je schovaná historická kontingencia: výber makromolekuly s istou špecifickou sekvenciou jednotiek (pričom takáto sekvencia nemusí byť jediná, môže byť degenerovaná) z množiny podobných makromolekúl: výber molekuly, ktorá má špecifickú funkciu, javí teda teleonómiu. V protiklade k jednoduchým atómovým alebo molekulárnym asociáciám, molekulárne rozpoznanie je *teleonomickým* procesom, výsledkom evolučnej kontingencie a výberu.

Molekula so schopnosťou sebakopírovania musí mať istú minimálnu komplexnosť; v tomto prípade sekvenčnú komplexnosť. Táto komplexnosť umožňuje molekule plniť funkciu, konať špecifickú prácu na prostredí, ktorej výsledkom je dynamické udržovanie *onticity* molekuly, jej pretrvávanie v prostredí. Aby tomu tak bolo, molekula musí mať určitú minimálnu znalosť o relevantných vlastnostiach prostredia, musí byť istá zhoda, akokoľvek hrubá a abstraktná, medzi týmito vlastnosťami prostredia a štruktúrou molekuly. Vo všeobecnosti, na všetkých úrovniach života, nielen na úrovni molekúl nukleových kyselín, komplexnosť, ktorá slúži určitej funkcii, a jedine ona, odpovedá *zabudovanej znalosti*, preloženej do konštrukcie systému. Prostredie je veľkou množinou možných ník: každá nika je problémom, ktorý treba rozriešiť, pretrvať v tejto nike znamená rozriešiť problém, a riešením je zabudovaná znalosť, algoritmus, predpis ako konať v záujme pretrvania.

[8] V tomto texte sa vychádza z bežnej teórie vzniku života, ktorá jeho počiatok vidí v objavení sa nukleovej kyseliny schopnej sebareplikácie. Je možná iná predstava súčasnej chemickej evolúcie proteínov, nukleových kyselín a lipidov; rozvinutá bola autorom a spolupracovníkmi v publikácii Kováč, L., Nosek, J., Tomáška, Ľ.: An overlooked riddle of life's origins: Energy-dependent nucleic acid unzipping. *J. Mol. Evol.* **57** (2003) S182-S189.

A tak je teda život od svojho počiatku poznávacím, kognitívnym systémom: sebakopírujúca molekula, sledujúca svoju onticitu vo svete, a teda najjednoduchší teleonomický systém, je už *subjektom* voči svetu, ktorý je objektom. Na všetkých úrovniach, od najjednoduchšej po najzložitejšiu, celková konštrukcia organizmu, inkarnácia uložených znalostí, prestavuje *epistemickú komplexnosť*. A práve epistemická komplexnosť kontinuálne rastie v biologickej evolúcii, rovnako ako v kultúrnej evolúcii, a dáva evolúcii jej nasmerovanie.

Čo nútilo pôvodné sebakopírujúce molekuly, aby zväčšovali svoju komplexnosť v priebehu evolúcie, aby sa asociovali s inými druhmi molekúl, aby vytvárali vyššie úrovne komplexnosti, a teda aj iné úrovne zabudovaných znalostí?

V pokusoch, ktoré robil Spiegelman so sebakopírujúcou nukleovou kyselinou fágu Q_β v umelom systéme, nedochádzalo k evolúcii k zväčšovaniu komplexnosti nukleových kyselín, ale naopak ku zmenšovaniu – najvyššiu darwinovskú fitnes mali molekuly, ktoré sa replikovali najrýchlejšie a tieto molekuly sa stávali kratšími, s menšou algoritmickou komplexnosťou [46]. Takto tomu muselo byť v jednoduchom a nemeniacom sa prostredí. V podobných novších pokusoch umelé evolučné reaktory fungujú v stacionárnom stave, s plynulým vtokom substrátov a výtokom produktov, ale tiež za jednoduchých a konštantných podmienok. Za takýchto podmienok sa evolučne selektujú jednoduché, rýchlo sa replikujúce molekuly [47].

Rozdielna je situácia v „evolučnom reaktore“ sveta. Replikujúce systémy sa nachádzajú v „nádobe“ neobmedzených rozmerov, v prostredí ktoré je komplexné a stále sa mení. Malá zmena prostredia stačí na to, aby sa molekula, ktorá sa rýchlo replikuje, ocitla v nepriaznivom prostredí, ktoré replikáciu spomalí a prípadne molekulu zničí. Systémy, ktoré sú *náhodou* komplexnejšie, a ktoré by boli v konštantnom prostredí eliminované, môžu naopak v takto zmenenom prostredí získať výhodu tým, že ich vyššia komplexnosť nesie v sebe potenciál pre „prežitie“ v nových podmienkach. Dalo by sa povedať, že ich väčšia komplexnosť predstavuje komplexnejší formálny systém, a teda väčšiu epistemickú kapacitu replikujúcej sa molekuly ako *subjektu*.

Čím je rýchlejšia replikácia komplexného systému, tým viac zdrojov systém spotrebováva, tým je väčšia disipácia voľnej energie v evolučnom reaktore a tým aj väčšia vzdialenosť od termodynamickej rovnováhy. Ku vzďaľovaniu sa od termodynamickej rovnováhy dochádza aj vtedy, keď sa začínajú vytvárať konzervatívne systémy, konštrukcie, nesúce viac zabudovaných znalostí. Aby sa konštrukcie udržali, už nie je potrebná disipácia energie: kinetické bariéry udržujú vzdialenosť od rovnováhy a spomaľujú prechod konštrukcií do rovnováhy.

Toto umožňuje, aby biologická evolúcia bola kontinuálnym rastom poznania: vytváraním subjektov so stále väčšími zabudovanými znalosťami, stále menej pravdepodobných, umiestnených stále ďalej od termodynamickej rovnováhy. Biologická evolúcia je *inventívna*, a práve tento fakt je príčinou, prečo je *progredujúca*.

Algoritmická komplexnosť sekvencie jednotiek v jednorozmernom priestore mohla spočiatku približne odpovedať komplexnosti sebareplikujúcej nukleovej kyseliny. Keď sa nukleové kyseliny neskôr asociovali s inými zariadeniami, membránami, tkanivami, jednotlivcami, spoločenstvami, komplexnosť takýchto systémov nadobúda podobu, ktorú už nemožno vyjadriť jednoducho ako sekvenciu ich nukleových kyselín. Aj komplexnosť samotných nukleových kyselín je oveľa väčšia ako je ich sekvenčná komplexnosť: táto komplexnosť zahrňuje aj časovanie prepisu génov. Takže k trom rozmerom priestoru

živého systému treba pridať čas ako jeho štvrtú dimenziu ([48] str. 89). Bude ukázané v inej publikácii, že táto komplexnosť má vzťah ku termodynamickej hĺbke Lloyda a Pagelsa [49]. A že k jej vyjadreniu je potrebné uvažovať subjekt a objekt ako nerozdeliteľnú jednotku, na čo v inej súvislosti upozornili Grassberger [50] a Gell-Mann [51].

Vzťah medzi epistemickou komplexnosťou, fitnes a pravdou je zložitý a bude analyzovaný v inej publikácii. Individuálny subjekt s veľkou epistemickou komplexnosťou javí robustnosť voči fluktuáciám prostredia. Ale aj veľký počet jednoduchých agentov, napríklad príslušníkov druhu, ktorý existuje vo veľkom počte kópií, rozmiestnených na veľkej ploche a úzko spolupracujúcich, môže byť robustný a prežiť rozmanité ataky z prostredia, ba aj v súťaži vytlačiť komplexné subjekty s rozsiahlym zabudovaným poznaním, pokiaľ je ich málo v porovnaní s jednoduchými agentmi. Práve táto zložitosť vzťahu slúži ako argument pre tých, čo popierajú ideu pokroku v evolúcii. Tento zložitý vzťah je zrejmý aj v kultúrnej evolúcii. Tvrdenie, že „väčšina má vždy pravdu“ je oveľa menej triviálne než by sa zdalo na prvý pohľad. Táto zložitosť robí epistemické „bludisko“ komplikovanejším, ale neprotirečí univerzálnej evolučnej tendencii.

Zabudované poznatky umožňujú teleonomickým systémom postupovať smerom k cieľom (spočívajúcim na ultimátnom cieli onticity) minimalizovaním dĺžky trajektórie k cieľu. Hľadanie cesty, spočiatku v podstate v podobe náhodného blúdenia, sa stáva v priebehu evolúcie stále menej randomným. V tomto zmysle je každý teleonomický systém racionálny, pričom stupeň racionality závisí od veľkosti zabudovaného poznania a je určený rozdielom medzi temer cikcakovitou čiarou a ideálnou trajektóriou v podobe priamky.

Biologická racionalita má jeden celkom podstatný nedostatok: je vždy racionalitou Russellovho moriaka (ktorý je rekurentnou témou Riedlovej epistemologickej analýzy, napr. [52]). Moriak, vykrmovaný sedliakom, očakáva budúcnosť ako pokračovanie súčasných výhod, bez toho, aby mal predstavu, že je to iba príprava na to, aby raz skončil na panvici. Anticipácia je prostou extrapoláciou z minulej skúsenosti. Jediná výnimka sa objavuje na úrovni racionality individuálnej ľudskej osoby: aj keď s veľkými obmedzeniami a ťažkosťami, človek má schopnosť obmedzeného predpovedania a robenia prognóz. No je to možno iba malinký záblesk v dejinách života na Zemi. Kultúrna evolúcia, s jej autonómnou dynamikou mémov, má možno racionalitu vyššiu ako má premýšľajúci ľudský jedinec, no pritom je to znovu len racionalita Russellovho moriaka.

A to je aj prípad vedy, ktorá býva označovaná za triumf ľudskej racionality. Ako už konštatoval Karl Popper [53], vedecká teória „môže preukázať svoju ‚fitnes‘ v podobe schopnosti prejsť úspešne tým testovaním, ktorým už prešla; lenže rovnako ako v prípade živého organizmu, ‚fitnes‘ znamená iba súčasné pretrvávanie a minulá úspešnosť nijako nezaručuje úspech v budúcnosti“.

Niet nijakých dôvodov, prečo by sebakopírujúce molekuly nevznikali aj inde vo vesmíre. Na základe rozumných predpokladov, vyplývajúcich z poznatkov súčasnej vedy, ľahko sa dá vypočítať z Drakeovej rovnice [54], že život musí byť bežným javom vesmíru. Vznik života zdá sa byť nevyhnutnosťou, nomickým procesom, a len po jeho vzniku nastupujú teleonomické procesy. Podľa Koperníkovho princípu [55] naša Zem, aj náš druh nemajú nejaké neobyčajné či výnimočné miesto v priestore a čase. Vesmír ako

celok vykazuje *epistemické rozvinutie* tým, že sa v ňom vyskytujú lokálne „ohniská“, v ktorých prebiehajú procesy hromadenia poznania. Metaforu bludiska, ktorou vysvetľujeme postup života na Zemi zároveň s neúspešnosťou ohromnej väčšiny aktérov nájsť cestu a neskončiť v slepej uličke, treba aplikovať na celý vesmír.

Tendencia k epistemickému rozvinutiu vesmíru je nazvaná epistemickým princípom. My, naša Zem, náš druh, sme aktérmi v tomto rozvinovaní. Majme stále na mysli: nezaberáme nijaké privilegované miesto vo vesmíre. Nemáme nijaký dôvod predpokladať, že z nespočetného množstva aktérov práve my sme vyvolenými. Ba máme hodne dôvodov pre domnienku, že tomu tak nie je.

2.5 Princíp račetovania

Ako už bolo konštatované, disipácia energie v evolúcii umožňuje nielen udržovanie disipatívnych štruktúr, ale aj vytváranie konzervatívnych štruktúr, konštrukcií. Konštrukcie sú systémy ďaleko od termodynamickej rovnováhy, oddelené od rovnováhy kinetickými bariérami. Termodynamika konštrukcií zatiaľ nebola rozpracovaná. Rast poznania v evolúcii znamená pribúdanie stále komplikovanejších konštrukcií.

Dynamika živých systémov na všetkých hierarchických úrovniach, spočíva v jednosmernom račetovaní[9]. Idea račetu bola do vedy zavedená Feynmanom, ktorý ju použil na ilustrovanie niektorých dôsledkov druhej vety termodynamiky; v podstate ako dôkaz, že rovnovážne fluktuácie nemožno použiť na konanie práce [56]. Jednoduchý mechanický račet je predstavovaný ozubeným kolieskom s asymetricky tvarovanými zúbkami a západkou, ktorá dovoľuje koliesku točiť sa len v jednom smere a bráni pohybu smerom dozadu. Na tomto jednoduchom princípe sú založené molekulárne, evolučné, vývinové, kognitívne, sociálne račety. Na molekulárnej úrovni konštrukcie dávajú životu možnosť využívať molekulárne račety: zariadenia, ktoré usmerňujú Brownov pohyb častíc v anizotropnom prostredí, v ktorom niet teplotných gradientov, síl a ani makroskopického elektrického poľa [57,58,59]. Randomný tepelný pohyb častíc je usmernený tak, aby slúžil funkcii, konaniu užitočnej práce na prostredí. Na prvý pohľad je tu zrejmý vzťah ku koncepcii racionality, ktorá bola načrtnutá v predchádzajúcom oddiele: náhodné blúdenie znamená „nijaká racionalita“, priamka odpovedá „ideálnej racionalite“ (ale tiež nevyhnutnému kauzálnemu spojeniu, *nomicite*) a čokoľvek medzi tým predstavuje „obmedzenú racionalitu“. Molekulárne račety sú zariadeniami čo vykazujú *molekulárnu racionalitu* a zároveň môžu byť označené za prototyp systémov s „ohraničenou racionalitou“.

Molekulárne račety boli navrhnuté najmä na vysvetlenie práce molekulárnych motorov, akými sú svalové proteíny alebo ATP syntéza. No v podstate aj jednoduché proteíny môžu fungovať ako račety. Platí to aj pre základné kognitívne zariadenia, molekulárne senzory. Brownovský pohyb molekuly liganda je usmernený elektrickým poľom senzora, ligand je priviazaný na senzor a časť energie viazania, miesto toho aby bola priamo disipovaná v podobe tepla, sa využije na vykonanie molekulárnej práce senzora: prenesenie signálu cez proteínovú molekulu. Viazanie receptora je spojené, *konjugované*, s prenosom signálu. Je to elementárna forma, na molekulárnej úrovni, fenoménu konjugácie. Podoby, počty, úrovne konjugácií narastali behom evolúcie. Život,

[9] Slovo račet je novotvarom, zavedeným ako ekvivalent anglického slova „ratchet“. V češtine sa občas používa slovo „rohatka“, technici na Slovensku a v Česku používajú aj slovo „račňa“.

popísaný týmito výrazmi, nie je nič iné než ohromný systém nespočetných konjugácií, ktorými je kanalizovaný tok energie, počínajúc radiačnou energiou zo Slnka. Energia preteká stupeň po stupni cez obmedzené, vybrané stupne voľnosti, až po konečný „výtokový kanál“ tepelnej energie, miesto toho aby bola disipovaná priamo. V tejto optike aj najzložitejšie ľudské činnosti, vrátane vedy a umenia, sú len evolučnými vsuvkami do nezastaviteľného toku a konečnej disipácie slnečnej energie.

Nepretržité vynaliezanie stále nových konštrukcií je samo procesom račetovania. Evolúcia ako celok je jeden ohromný račet. V evolučnej biológii je dobre známy Mullerov račet: hromadenie škodlivých mutácií, ktorým narastá mutačná záťaž, a ktoré nevyhnutne končí vyhynutím ([60] str. 241). Mullerov račet je len virtuálny, len výnimočne naozaj funguje: sex bol vynálezom, ktorým sa život bráni Mullerovmu račetu. Tento obranný účinok sexu je zosilnený račetami outbreedingu, predstavovanými rôznymi mechanizmami, ktoré zabraňujú príbuzenskému kríženiu, vrátane *tabuizovania incestu* u ľudí.

Dobrou ilustráciou najelementárnejších evolučných račetov je počítačové modelovanie evolúcie RNA [61]. Pravdepodobnosť prechodu medzi dvomi rozdielnymi molekulami RNA, ktoré sa od seba odlišujú jedinou mutáciou, nie je symetrická: deštrukcia štruktúrneho elementu jednobodovou mutáciou je ľahšia než jeho vytvorenie.

Najúčinnejší evolučný račet predstavuje Weismannova bariéra, ktorá oddeľuje genotyp a fenotyp. Vďaka Weismannovmu račetu je žitie individuálneho organizmu, so všetkými jeho mýlením sa, sklamaniami, degeneráciou, stareckou rezignáciou, oddelené od ihriska *evolučného hádzania kockou*, ktoré prebieha na úrovni genómu, a ktoré – ak neberieme do úvahy prípady parazitizmu – posúva genóm jednosmerne ku stále väčšej komplexnosti. To z tejto elementárne molekulárnej podoby vyrastá Campbellov jav „variácií a selektívnej retencie“, ktorý pozorujeme na všetkých úrovniach evolúcie, biologickej i kultúrnej [27].

Už bola zmienka o tom, že evolúcia ťahá organizmy stále ďalej od termodynamickej rovnováhy. Prigogine so spolupracovníkmi opakovane zdôrazňoval (napr. [62]), že hmota ďaleko od rovnováhy je tvorivá. Čím mohutnejšia je disipácia energie, tým účinnejšia je sebaorganizácia a tým viac poriadku sa vytvára ireverzibilným spôsobom. Oceňujúc pionierske štúdie Prigogina, môžeme túto celkovú tvorivú činnosť evolúcie nazvať Prigoginovým račetom.

Račetovanie hrá dôležitú rolu v individuálnom vývine[10]. Jednoduchou analógiou so skvelým objavom Georgea Beadlea z ranného obdobia biochemickej genetiky, že vo všeobecnosti jeden gén kóduje jeden enzým, ktorý dostal názov „hypotéza jeden gén-jeden enzým“, to isté by sa mohlo povedať o morfogenéze: „jeden gén-jeden morfologický znak“. Hoci pravdepodobne nijaký vedec takúto analógiu nenavrhol, slúžila ona ako výzva k útokom zo strany rôznych anti-darwinistov a biologických štrukturalistov. Keby každý diskrétny znak mnohobunkového organizmu mal byť určený samostatným génom, a to na každom jednotlivom stupni vývinu, organizmy by museli byť vybavené nejakou zázračnou „superkomplexnosťou“. Dokonca možno aj predstava genómu nie ako plánu, ale ako receptu na upečenie koláča, ktorú navrhol Dawkins ([63] str. 294), hoci predpokladá menšiu zložitosť, je stále možno ešte prináročná. Je preto možné, že pri vývine je v činnosti nejaký druh somatického darwinizmu [64,65,66], ktorý je ale obmedzený takým spôsobom, že vždy vyúsťuje do jedného, kvázi-determinovaného stavu,

[10] Treba odlišovať slovo „vývoj“, ktorým sa prekladá anglické slovo „evolution“ a slovo „vývin“, čo je prekladom slova „development“.

ale bez toho, že by vyžadoval veľmi náročný program alebo nejakú ústrednú autoritu, „genómový mozog“ (či nebodaj „genómový ústredný výbor“). Na každom stupni morfogenézy, počínajúc dvomi bunkami najrannejšieho embrya, môže dochádzať ku kombináciám rôznych čisto lokálnych asociácií, v podstate stochastických alebo len málo usmernených, ale činnosťou vývinových račetov špecifických pre každý vývinový stupeň, na každom stupni by sa zachovali len vhodné a tým dávali možnosť pre ďalšie asociácie na vyššej úrovni. Komplexný globálny poriadok by sa takýmto spôsobom spontánne vynáral z výlučne lokálnych interakcií jednoduchých jednotiek.

Toto sa javí ako univerzálny princíp usporadúvania v evolúcii, vývine, kognícii, ba i štruktúrovaní megaspoločenstiev. Princíp minimálnej komplexnosti naznačuje, že by práve molekulárna biológia mohla urobiť podstatný príspevok k jeho vyjasneniu. V našej dobe dostáva tento princíp najväčšiu podporu zo štúdií umelej inteligencie [67]: komplikované správanie sa vynára z interakcií jednoduchých autonómnych agentov, ktorí sú situovaní vo svojom prostredí, pričom každý sleduje len jednoduchý cieľ. Niet tam nijakého ústredného plánovača, ani vnútornej štruktúry, ktorá by predstavovala „plán“ systému, ani neexistuje centrálna reprezentácia, ktorú by zdieľali všetky agenty.

Princípu račetovania je implicitný iný princíp. Možno ho nazvať princípom *hierarchickej kontinuity dizajnu* a voľne ho spojiť s Jacobovým princípom zmajstrovávania, brikoláže (tinkering, bricolage) [68]. Evolučné hádzanie kockou, skombinované s evolučným račetovaním, robí z evolúcie proces, v ktorom niečo nové vždy nutne stavia na tom, čo už bolo dosiahnuté. Zariadenia a konštrukcie, vynájdené v minulosti, sú pozmenené do novej podoby. Dokonca aj keď nie sú funkčné, zriedkakedy sú celkom odhodené. Skôr fungujú ako rezervy, v celku alebo rozmontované na svoje súčiastky, ktoré sa dajú využiť ako moduly v iných kombináciách a pre iné účely. Toto je aj jedna z príčin narastajúcej redundancie v evolúcii na všetkých úrovniach biologickej organizácie, od redundancie génov až po veľkolepú redundanciu neurónov v ľudskom mozgu a v jej dôsledku aj ľudskej kultúry. Zvyšovaním redundancie v evolúcii sa život menil z bojovnej hry o prežitie na hru pre zábavu (z *game* na *play*). Adaptívne správanie sa doplňovalo *expresívnym správaním*. Po nástupe kultúrnej evolúcie sa k týmto dvom formám správania pridala tretia : *správanie vynútené mémami*.

Kontinuita dizajnu, ktorá tvaruje a obmedzuje evolúciu, je fundamentálne dôležitá vo vývine. Vo vývine ešte prísnejšie než v evolúcii každý nový úkon nadväzuje na úkon predchádzajúci. Keďže vývinové úkony sa neustále vetvia, v celej hierarchii úkonov výsledok každého predchádzajúceho je zatláčaný stále hlbšie. Keďže vývin je geneticky riadený, výsledky najrannejších úkonov, nachádzajúce sa hlboko v základoch, sú temer nezmeniteľné. Akákoľvek zmena v ranných génoch embryonálnej fázy, bez ohľadu na to, ako by mohla byť výhodná, predpokladá súčasné znovuupravenie všetkých génov, čo určujú neskoršie fázy ontogenézy – pravdepodobnosť takého simultánneho uspôsobenia je prakticky nulová. V evolúcii vývinu dochádza k postupnému nezvratnému zafixovaniu všetkých predchádzajúcich úkonov. Wimsatt [69] nazval tento proces *generatívnym vnáraním* (generative entrenchment). Jeho faktickú ireverzibilitu ilustroval modelom vývinového zámku: je to valcový zámok s krúžkami, pričom každý krúžok môže mať viac nastavení, no nastavenie krúžku závisí od nastavenia predchádzajúceho krúžku. Zmena v nastavení krúžku, ktorý je na začiatku procesu, si vynúti zmenu nastavenia všetkých ostatných, čo ale opačne neplatí. Vlastne takýto digitálny vývinový zámok možno vidieť ako druh vývinového račetu. Dobrým konkrétnym biologickým príkladom, na ktorom je

zrejmá ako kontinuita dizajnu tak zmajstrovávanie, je gén *Pax6* [70]. Gén, ktorého počiatočnou funkciou bolo riadenie diferenciácie fotoreceptora v primitívnom „oku“, ktoré pozostávalo iba z fotoreceptorov, bol evolúciou posunutý do funkcie hlavného regulačného génu v oku hmyzu, myši i človeka. Riadi gény, ktoré boli pridané v evolúcii neskôr, ako sa oko stávalo stále komplikovanejším.

Na tom istom všeobecnom princípe spočíva ľudská kognitívna ontogenéza. Blízko pri dne leží *vpečaťovanie* (imprinting). Za tým časovo nasleduje prvotné naplnenie geneticky určených abstraktných, ale pritom špecifických „poštových schránok“ mysle konkrétnymi pojmami, ideami a návykmi – *incipácia*. Do hierarchie sa potom radí spevňovanie kontingenciami prostredia a nakoniec vrcholí vedomé uvažovanie. Na všetkých týchto úrovniach bežia variácie, lokálne interakcie, selektívna retencia, nové úrovne sa budujú nad neodstrániteľnými nižšími a výsledkom je zjednotená, koherentná a dynamická štruktúra. Vo všeobecnosti čím včaššie v individuálnom živote sa kognitívny modul zmontuje, tým rezistentnejším je k neskoršiemu pozmeňovaniu. Husserlov termín *Lebenswelt*, ktorý od neho prevzali existencialisti, ktorý je jedinečný každému ľudskému jedincovi a zjavne neprístupný iným, označuje realitu, ktorá je takýmto jedinečným spôsobom konštruovaná v ontogenéze každej jednotlivej osoby [71]. A vlastne aj vznik a vývoj každej špecifickej ľudskej kultúry sa riadi tým istým princípom [72].

Račetovanie v evolúcii, vo vývine, v kognícii plní tú istú funkciu: zaisťuje stupňovité hromadenie a zmysluplné využitie poznania a znemožňuje jeho zmenšovanie alebo rozpad tým, že by procesy bežali smerom dozadu. Račety bežia na všetkých úrovniach, od molekúl po megaspoločenstvá. Pri analýze hierarchie týchto úrovní je užitočný pojem *granulácie*, ktorý sa vyskytuje v štatistickej fyzike. Jeho významom sa bude zaoberať iná publikácia.

2.6 Princíp minimálneho predsudku

„Od svojho samého začiatku musí byť život vybavený všeobecným poznaním, poznaním, ktoré obyčajne nazývame poznaním prírodných zákonov. Samozrejme nie v zmysle vedomého poznania“ ([53] str. 32). Toto vyjadrenie Karla Poppera je nešťastné. Je možno jednou z príčin, prečo Popper vznik života považoval za „neuveriteľne nepravdepodobnú“ udalosť. Život je *obmedzený všetkými* prírodnými zákonmi, neznamená to však, že ich všetky musel od začiatku (hoci aj nevedome) poznať. Sám Popper, ako všetci evoluční epistemológovia, tvrdil, že poznanie prostredia znamená *adaptáciu* naň („...die Anpassung des Lebens an seine Umgebung ist eine Art von Erkenntnis“). Z toho vyplýva, že adaptácia na jednoduché prostredie sa rovná jednoduchému poznaniu, a že behom evolúcie trvalý rast zložitosti ník, ktoré život obsadzuje, znamená trvalý rast poznania – teda, trvalý rast epistemickej komplexnosti.

Obmedzené poznanie bolo významnou príčinou vyhynutia, extinkcie druhov. Prírodné zákony platia, aj keď ich život nepozná. Pokiaľ ide o pravdepodobnosť vyhynutia, iná charakteristika živých bytostí, neodkloniteľnosť ich presvedčení, fanatizmus, už nie je tak jednoznačná. Organizmy sú *fanatici*. Jednoduché organizmy, ktoré nemajú schopnosť učenia sa, ba aj jednoduché samokopírujúce nukleové kyseliny, sú absolútnymi fanatikmi. Organizmy nevynaliezajú a neudržujú si hypotézy, ale oplývajú *vierami* a iba niektoré z týchto vier odpovedajú, v špecifickom prostredí, poznatkom o jeho vlastnostiach. Čo znamená, že len niektoré konštrukcie organizmov predstavujú zabudované poznanie; iné predstavujú *zabudované viery*. Akonáhle je raz adaptovaný na

svoje prostredie, jednoduchý organizmus je úplne nemenný. Ak by sme novú mutáciu u baktérie považovali za novú vieru o prostredí, môžeme povedať, že mutant obetuje svoj život, aby dokázal vernosť svojej viere. U organizmov, ktoré majú schopnosť učenia, situácia nie je natoľko rozdielna, ako by sme predpokladali: ako už bolo uvedené, kontinuita dizajnu sa stará o to, aby tie viery, ktoré boli získané v rannej ontogenéze, ostávali zachované a nemenné behom celého individuálneho života. Z hľadiska spoločného génového poolu je to zariadenie adaptívne: je to jeden z hlavných zdrojov vytvárania a udržovania polymorfizmu, ktorý slúži prežitiu druhu vo fluktuujúcom prostredí. Tým, že sa takto zvyšuje variancia, zväčšuje sa evolučný potenciál druhu. Pri jednej konkrétnej zmene, fluktuácii prostredia, niektorí fanatici zahynú, kým iní, s inými vierami, zväčšia svoju fitnes; keď sa evolučné kyvadlo vychýli v opačnom smere, šance pre obe kategórie fanatikov sa obrátia. No druhy, čo by boli zložené iba z liberálov, iba zo skeptikov, iba z oportunistov by neboli dostatočne robustné, aby prežili fluktuácie a nevyhynuli.

Ľudský druh nie je výnimkou. Človek, podobne ako iné živočíchy, nie je popperovským racionalistom, ktorý by horlivo vystavoval svoje výklady sveta testovaniu a bol ochotný nahradiť ich novými. Ľudia sú *mýtofili*: lipnú pevne na svojich vierach, často ochotní obetovať za ne vlastné životy rovnako zaťato ako baktérie.

Prostredie, v ktorom sa formovali prirodzeným výberom naše mentálne schopnosti, nepodporovalo fitnes jednotlivcov, čo boli nerozhodní, váhajúci, tolerantní, ohromení zložitosťou sveta a náchylní ku kognitívnym zmätkom a existenciálnej úzkosti. V evolúcii musel byť aj silný selekčný tlak na sociálnu konformitu. Mýty, nezmeniteľné, nediskutovateľné, vševedúce, všetko zahrňujúce vysvetlenia sveta, boli najúčinnejším skupinovým „lepidlom“, ktoré robilo zo sociálnej skupiny účinnú kohéznu jednotku a naopak podnecovalo súťaž medzi skupinami. Mýty plnia svoju funkciu aj dnes vo forme organizovaných náboženstiev a vo forme ideológií. „Nedostatky ľudského rozumu“ [28], ktoré vedú ku kognitívnym ilúziám a ku krajným predsudkom, fungujú ako vynikajúca živná pôda pre zdarné množenie a šírenie mémov, ktoré predstavujú kolektívne mýty a ideológie. Pre spevnenie skupinovej kohézie je zvlášť dôležitou kognitívnou ilúziou individuálny, ale aj skupinový *sebaklam*. Sokratova výzva „poznaj sám seba“ je oveľa ťažšie splniteľná než sme pri svojom sebaklame schopní pripustiť.

Mnohé črty fyzickej a mentálnej výbavy ľudí, ktoré boli vyselektované v malých neanonymných skupinách lovcov a zberačiek v savane, nie sú viac adaptívne v sociosvete, ktorý vytvorila kultúrna evolúcia. Aspoň čiastočné korekcie môže do tejto diskrepancie zavádzať jedinečný objav európskej kultúry: experimentálna veda. Nie ani tak vedomou činnosťou vedcov – jednotlivý vedec nemusí byť o nič menej mýtofilný ako laik, tiež sa pevne drží svojich vier – ale spôsobom, akým veda prekladá svet do podoby reality: experimentálne výsledky nepripúšťajú odchýlenia od racionálnosti, musia byť v súhlase s prírodnými zákonmi, z ktorých možno väčšina je vyjadrením nejakých extremálnych princípov. Experimentálny prístup ukladá na prácu vedcov princíp, ktorý sám tiež je extremálnym princípom, a ktorý nie je celkom vyhovuje „prirodzeným“ mentálnym sklonom človeka: princíp minimálneho predsudku. Možno ho nazývať aj Jaynesovým princípom, podľa fyzika, ktorý prvý tento princíp presne formuloval. Princíp sám bol implicitný v názoroch niektorých filozofov, napr. v metafore Occamovej britvy ([73] str. 462), v požiadavke ekonómie myslenia [74], alebo v zásade parsimónie [75].

Jaynes princíp matematicky formuloval. Ak má niekto neúplnú znalosť o niečom, pripísanie minimálne predpojatých pravdepodobností je také, čo pri daných dátach maximalizuje Shannonovu informačnú entropiu [76]. Jaynes elegantne ukázal, že niektoré teorémy termodynamiky sa dajú vyvodiť ako dôsledky tohto princípu. A hádam nie je ani prehnané očakávať, že aj základy niektorých iných vied by sa dali odvodiť z Jayensovho princípu.

Zrejmé je prepojenie medzi princípom minimálneho predsudku a tým chápaním racionality, aké bolo popísané vyššie. Paradoxne však práve úvahy o racionalite sa často nedržia tohto princípu. Racionalita sa chybne chápe ako záležitosť čisto rozumového zdôvodňovania, mentálneho výpočtu, vedomého zvažovania ziskov a strát. Neberie sa do úvahy „*múdrosť tela*“, ktorá sa dosiahla v evolúcii prirodzeným výberom, možnosť racionálneho konania sa upiera iným živočíchom než je človek, racionalita sa pripisuje len individuálnej ľudskej osobe a zabúda sa na racionalitu adaptívnych dynamických systémov. A nadovšetko sa neuznáva racionalita emócií a ich rozhodujúca rola v zmysluplnom správaní. Viera na silu individuálneho ľudského rozumu bola veľkým predsudkom, najmä v európskej kultúre. To je predsudok, ktorý nemá nijaké evolučné opodstatnenie. Práve princíp minimálneho predsudku opodstatňuje ďalší princíp kognitívnej biológie.

2.7 Princíp minimalizovania utrpenia

Keby baktérie mali vedomie a schopnosť sebareflexie, ich svetonázor by celkom isto bol „baktériocentrický“. To isté platí pre potkanov, až na to, že ich konštrukcia reality by bola „ratocentrická“. Jedni i druhí by sa hrdili svojou evolučnou úspešnosťou a pozerali by sa s pohŕdaním na človeka, ktorého charakterizujú také evolučné podivnosti, privodené redundanciou jeho mozgu a s ňou spojenou kultúrnou evolúciou, ako je atómová bomba alebo masová televízna zábava. My, ľudia, si môžeme postaviť teoretickú otázku, ako to asi je byť netopierom [77], ale nikdy nebudeme schopní preniknúť do druhovo-špecifickej reality netopierov. Obmedzení sme na našu realitu špecifickú pre človeka a svojimi evolučnými a vývinovými račetmi sme nútení byť antropocentrickí.

Bakteriálny filozof, odsúdený na doživotné pobývanie vo svojom druhovo-špecifickom formálnom systéme, by si staval tie isté dve najvšeobecnejšie otázky, aké si po dva a pol tisíc rokov stavali ľudskí filozofi: (1) Existuje svet alebo je len mojou ilúziou? (2) Ak niečo existuje, prečo to existuje, prečo je vôbec niečo a prečo nie je jednoducho nič? Tak ako jeho ľudský kolega, ani bakteriálny filozof by na otázku nenašiel odpoveď. Aby ju našli, museli by, jeden i druhý, vykročiť zo svojich formálnych systémov – a to je nemožné.

Evidenciu, a to naprosto nezvratnú, získavame my ľudia iným spôsobom. Získavame ju *vedomým prežívaním emócií.* Tým, že si uvedomujeme radosť a bolesť.

Emócie sa vyvinuli ako adaptívne zariadenie, ktoré účinne zaisťuje onticitu živých bytostí. Bolo urobených veľa pokusov vystopovať ich evolučný pôvod [78,79,80]. Tak, ako v prípade kognície, názory sa líšia a ležia medzi dvomi extrémami: pri jednom extréme sa emócie považujú za výlučne ľudské privilégium, druhý extrém pripisuje emócie nielen všetkým živým útvarom, ale aj niektorým neživým objektom, napr. termostatu. Užitočným sa môže ukázať použitie princípu minimálnej komplexnosti. Takto možno hľadať emócie v akomkoľvek teleonomickom systéme, v ktorom sú zabudované zariadenia na vyhodnocovanie stimulov z prostredia [81].

U druhu *Homo sapiens* sa pozoruje zvláštnosť, a možno dokonca jedinečnosť: koincidencia emócií a sebauvedomovania. Práve paralelizmus, či skôr neoddeliteľnosť týchto dvoch evolučných výdobytkov, dávajú človeku možnosť dvojitej transcendencie – vykročiť z formálneho systému logického zdôvodňovania a prekonať fundamentálny imperatív života: úsilie o onticitu, a o nič iné než o individuálnu onticitu, za každú cenu! Neskôr budú uvedené argumenty pre názor, že sebauvedomenie a schopnosť silných emócií, vlastne priam hyperemocionálnosť, sú konečnými produktmi jediného *evolučného trieľu*, evolučnej eskalácie (run-away), takže oba predstavujú viac než adaptívny výdobytok evolúcie. Emócie sú hlavným motorom expresívneho správania[11].

Napriek úspešným pokusom vysvetliť u človeka altruizmus tými istými elegantnými zdôvodneniami, aké zrejme platia pre všetky biologické druhy, vrátane rastlín a mikroorganizmov, vieme všetci z introspekcie, že ľudský altruizmus má aj vlastný, veľmi účinný zdroj: *empatiu*. Empatia nielen riadi správanie ľudského jedinca voči iným ľudským bytostiam; poskytuje mu aj najprenikavejšiu evidenciu o tom, že nielen on, ale aj iní naozaj existujú [82]. Je to presvedčivejší argument, než by sa získal len racionálnym bayesovským zdôvodňovaním. Bayesovské zdôvodnenie (na ktorom spočíva princíp najmenšieho predsudku) opodstatňuje iný argument, Koperníkov princíp: nikto z nás nezaberá nijaké mimoriadne, privilegované miesto vo vesmíre. Oba argumenty vedú k záveru, ktorý už pred tromi dekádami formuloval Linus Pauling [83]:

„*Evidencia mojich zmyslov mi hovorí, že som človek taký, ako iní ľudia. Keď sa porežem, som poranený, trpím, nariekam. Vidím, že aj iný človek, keď sa poreže, narieka. Z jeho správania robím záver, že aj on trpí rovnako, ako ja. Nič z mojich pozorovaní mi nedáva dôvod veriť, že som niečím špeciálnym, čo by ma zásadným spôsobom odlišovalo od iných ľudských bytostí: miesto toho som vedený k presvedčeniu, že som človek taký ako ostatní. Chcem byť oslobodený od utrpenia ako sa len najviac dá. Chcem žiť šťastným a užitočným životom, spokojným životom. Chcem, aby mi iní ľudia pomohli byť šťastným, aby mi pomohli udržovať moje utrpenie na minime. Je potom mojou povinnosťou pomôcť zas im, aby boli šťastní, snažiť sa o zabránenie utrpenia iných ľudí. Tento argument ma vedie k fundamentálnemu etickému princípu: rozhodovania sa medzi alternatívami konania musia byť robené tak, aby sa minimalizovalo ľudské utrpenie...*

Tvrdím, že princíp minimalizovania ľudského utrpenia je vedeckým princípom, s logickým, vedeckým základom. Súhlasím s profesorom Jeacquesom Monodom, ktorý povedal, že etika musí spočívať na axiómach, rovnako ako geometria. Profesor Waddington poukázal na to, že hoci je možno rozvinúť geometriu na báze iných axióm, všetci ľudia súhlasia s tým, že v praktickom svete treba prijať Euklidove axiómy. Mám pocit, že aj keď máme teoretickú slobodu, aby sme formulovali rozmanité etické systémy, výber rozumného a praktického etického systému je silno obmedzený našimi znalosťami o povahe fyzikálneho a biologického sveta, a že jedinými prijateľnými etickými systémami sú tie, čo sú založené na princípe minimalizovania ľudského utrpenia.“

Etická norma sa môže udržovať v populácii dvoma spôsobmi. Môže pretrvávať tým, že zvyšuje fitnes jedincov, ktorí sa ňou riadia. V najjednoduchšom prípade by to bol

[11] Objavenie sa emócií a sebauvedomovania ako výsledok evolučného trieľu bolo medzitým analyzované v publikácii *Ľudské vedomie je produktom evolučnej eskalácie emocionálneho výberu*. V: Sinčák, P., Kvasnička, V., Pospíchal, J., Kelemen, J., Návrat P. (eds.): Slovensko-České rozpravy o umelej inteligencii. Stará Lesná: Slovenská spoločnosť pre umelú inteligenciu 2003, str.21-39. Expresivita a expresívne správanie budú popísané v inej publikácii.

dôsledok recipročného altruizmu – akéhosi výpočtu, ktorý je vlastne implicitný aj v takých univerzálnych morálnych prikázaniach ako je „nerob svojmu blížnemu to, čo nechceš, aby on robil tebe“. Ale môže sa tiež jednať o silno virulentný mém, ktorý sa šíri populáciou aj vtedy, ak znižuje fitnes tých jedincov, čo sú ním infikovaní. Princíp minimalizácie utrpenia, založený na empatii, kombinuje biologický i kultúrny príspevok. Empatia ako biologická črta by bola jedným z tých abstraktných mentálnych „poštových schránok“, o ktorých už bola zmienka, ktorá môže byť zaplnená buď súcitom, charitou a obetavosťou, alebo závisťou, pomstivosťou či zlomyseľnou krutosťou. Ani jeden z týchto dvoch typov prejavov správania nepozorujeme u iných živočíchov než je človek, čo svedčí o tom, že ich dovoľuje len vlastnosť špecifická pre človeka. Záleží len na konkrétnom zložení mémového poolu v nejakom kultúrnom prostredí, ktorý z dvoch typov správania prevláda.

Príroda je ľahostajná k utrpeniu. Časť ľudského utrpenia je výsledok úmyselného konania ľudí, ale asi väčšia časť je nezamýšľaná, spôsobená neznalosťou a bezmocnosťou tvárou v tvár slepým prírodným a spoločenským silám. Nevedomosť, neschopnosť, pocit menejcennosti, strach predstavujú živnú pôdu pre mémy závisti a krutosti, ich opak podporuje zhovievavosť, súcit, oddanosť, spoločenskú hravosť. Po celé storočia ľudia vedu vysvetľovali a opodstatňovali ako činnosť zameranú na zmenšovanie neznalosti a bezmocnosti. Všetky princípy kognitívnej biológie opodstatňujú takéto chápanie vedy ako organizovaného zmenšovania neznalosti. Na druhej strane však reťazec tu uvedených argumentov spojuje vedu priamejšie aj so zmenšovaním utrpenia. „Hľadanie pravdy“ sa často považovalo za vnútornú normu vedy. Nie je však tomu tak: veda, ktorá klame, jednoducho vedou nie je. Princíp minimalizovania utrpenia pridáva vede ďalšiu dimenziu. Zas nie ako normu: čím je veda „opravdivejšou“, tým je bližšia tomuto extremálnemu princípu. Kým Paulingova argumentácia končila normatívnym návrhom ako základom etického systému, toto konštatovanie nie je normatívne, ale len popisné. Vznik vedy a jej ďalšia evolúcia ako inštitúcie boli samou svojou podstatou zviazané so zmenšovaním ľudských trápení: bolesti, ťažkej práce, biedy, úzkosti. Veda sa stala hlavným nástrojom v ľudskom snažení o minimalizovanie bolesti a maximalizovanie príjemnosti. Kognitívna biológia len vysvetľuje, prečo je tomu tak.

Z toho nevyplýva, že vedecké bádanie, ktorým sa spôsobuje utrpenie, či už nezamýšľané alebo aj zamýšľané, nie je vedou. Je vedou, ale so značným podielom neznalosti. Tak ako život na Zemi, či život v celom vesmíre, aj veda postupuje dopredu v bludisku: je v ňom veľa chybných ciest a slepých uličiek, ale cesta vpred v ňom predstavuje hlavnú evolučnú tendenciu. Úspech však nie je vopred predpísaný.

3 Diskusia

Potrebujeme nový termín kognitívna biológia? Je kognitívna biológia súčasťou evolučnej epistemológie alebo je totožná s evolučnou epistemológiou a len sa skrýva za iný názov? Skôr platí niečo opačné: kognitívna biológia je širšou množinou s evolučnou epistemológiou ako podmnožinou. Kognitívna biológia vyrástla z molekulárnej biológie, s predpokladom, že vyjasnenie molekulárneho rozpoznávania, spracovávania molekulárnych signálov, organizácie génových sietí, proteínového počítania by mohlo poskytnúť kľúč k pochopeniu vyšších kognitívnych procesov. Zároveň jej úzke spojenie s fyzikou a chémiou by mohlo pomôcť ukončiť neustále spory a zmätky okolo vzťahu medzi

informáciou a fyzikálnou entropiou a umožniť presnejší a formalizovanejší popis poznania, epistemickej komplexnosti a racionality.

Niektorí významní fyzici naznačili, že extremálne princípy fyziky by mohli mať fundamentálny význam pre naše pochopenie prírody ([84,23,85]. V zmysle logickej parity by to mohlo platiť aj pre pochopenie fungovania mysle. Hádam nie náhodou sa niektoré princípy kognitívnej biológie podobajú, alebo majú vzťah, k extremálnym princípom fyziky. Možno sa raz ukáže, že duálnosť ontológie a epistemológie je iba kognitívnou ilúziou.

Bol vyslovený názor, že teória vedeckých revolúcií Thomasa Kuhna, s jej koncepciou zmeny paradigiem, mala zlý vplyv na kognitívne vedy: mnohí kognitívni vedci predstavujú svoje teórie ako nové paradigmy, popierajúce iné [86]. Kognitívna biológia nepredstavuje nijakú zmenu paradigmy. Stavia na tradičných pohľadoch viacerých vedeckých disciplín, len sa usiluje o syntézu. Aj keď sa to možno na prvý pohľad nezdá, treba zdôrazniť, že myšlienka o centrálnej role ľudského vedomého prežívania emócií čerpá z hlbokých zdrojov fyziky a molekulárnej biológie. Naznačuje pritom dôsledky tohto prežívania, ktoré možno vyjadrujú revolučný zlom v biologickej evolúcii: už nie viac onticita sebeckých génov, permanencia sebareplikujúcich entít, ale utrpenie, ktoré vedome prežíva živý jedinec, sa stalo – aj keď možno len na krátku dobu a možno ani nie po prvý raz v dejinách vesmíru – stredom zápletky v dráme, ktorá sa odohráva na javisku sveta.

Treba zdôrazniť, že žiaden z princípov kognitívnej biológie, dokonca ani princíp minimalizácie utrpenia, nie je normatívnym. Sú to popisné konštatovania odvodené z ich axióm. Bolo by predčasné pokúsiť sa o axiomatizovanie kognitívnej biológie. Dostatočným dôkazom čisto popisnej povahy týchto princípov sú niektoré základné postuláty, vyjadrené v podobne idealizovanej podobe v akej sú vyjadrené postuláty Euklidovej geometrie. Prvým východzím postulátom je existencia, onticita ľudského utrpenia. Druhým postulátom je empatia ako iná emocionálna kvalita. Minimalizácia svojho vlastného utrpenia je, v ideálnom prípade neobmedzenej empatie, neoddeliteľná od minimalizácie utrpenia blížneho. Veda ako špecifický ľudský vynález je v tejto činnosti nápomocná. Netreba sa pritom odvolávať ani na individuálnu či inkluzívnu fitnes, ani na recipročný altruizmus. Tvrdenie, že tento princíp minimalizácie utrpenia je normatívny, by bolo ekvivalentné tvrdeniu, že extremálne princípy fyziky, napríklad princíp najmenšieho účinku, sú normatívne.

To, samozrejme, neznamená, že vedci sú vo svojej činnosti poháňaní súcitom a vedomou snahou byť užitočnými pre ľudstvo. Tendencia vedy zmenšovať ľudské utrpenie plynie z jej vnútornej dynamiky, v ktorej prijímanie a podpora vedy, ale aj technovedy verejnosťou hrá podstatnú rolu. Trend k úplnému odstráneniu utrpenia, k jeho redukcii na nulu, ktorý je trojjediným dôsledkom evolučnej eskalácie ľudskej hyperemocionality, vedomia a mémovej dynamiky (ktorá v sebe zahrňuje aj vedu), prekračuje ľudskú biológiu a to aj ľudskú darwinovskú fitnes: utrpenie malo adaptívnu povahu v úsilí o existenciu; a nulové utrpenie (v kombinácii s maximom umelých príjemností, ktoré možno budú zanedlho poskytovať techniky virtuálnej reality) môže aj toto úsilie o existenciu zredukovať na nulu. Lenže musí byť prežitie naďalej najvyššou hodnotou pre ten druh vedomého života, aký sa vyvinul na Zemi?

Toto konštatovanie je zásadné pre porozumenie rozsahu toho, o čo usiluje kognitívna biológia. Bolo by omylom chápať ju iba ako pokus o vyjasnenie molekulárnych

základov kognície. Princíp račetovania, v kombinácii s epistemickým princípom, vysvetľuje, ako jednosmerný beh evolučných a vývinových račetov generuje stále väčšiu komplexnosť, ktorá teraz vyvrcholila v podobe ľudskej vedomej emocionality a v podobe vedy. Kognitívna biológia zahrňuje ľudskú transcendenciu, ale zároveň usiluje o postihnutie jej biologických, a molekulárnych, koreňov.

Kognitívna biológia je nielen vedou, ale aj konceptuálnym programom. Hoci nie ako veda, ako program má aj normatívne aspekty. Program má svoje opodstatnenie v presvedčení, ktoré možno po prvý raz vyjadril Claude Lévi-Strauss: 21. storočie bude storočím vedy o človeku – alebo nebude. Hlavné konštatovanie programu znie takto: Cieľom vedy *musí byť* minimalizovanie ľudského utrpenia *plus* optimalizácia ľudskej hmotnej a duchovnej pohody. (Toto konštatovanie, na rozdiel od predchádzajúceho popisného konštatovania o zjavnosti tejto tendencie v evolúcii vedy, je normatívne.) Kognitívna biológia ako program sa pritom drží zásady, že pre plnenie tohto príkazu je pri súčasnom stave poznania treba študovať ktorúkoľvek z čŕt ľudskej prirodzenosti za využitia organizmov s minimálnou komplexnosťou, ba aj čisto molekulárnych systémov, u ktorých sa ešte táto črta vyskytuje. Vstupné motto tejto štúdie možno trochu modifikovať (a vyjadriť s mierne normatívnou konotáciou): najúčinnejší prístup k štúdiu človeka, k zmenšovaniu utrpenia a zaisťovaniu ľudského šťastia bude možno ešte po nejakú dobu spočívať v štúdiu baktérií.

Literatúra

[1] Koshland Jr., D. E.: A response regulator model in a simple sensory system. *Science* **196** (1977) 1055-1063.

[2] Lorenz, K.: *Der Abbau des Menschlichen*. München: Piper 1983.

[3] Butler, D.: Advances in neuroscience „may threaten human rights“. *Nature* **391** (1998) 316.

[4] Piaget, J.: *Biologie et connaissance*. Paris: Gallimard 1967.

[5] Radnitzky G., Bartley III, W. W. (eds.): *Evolutionary epistemology, rationality, and the sociology of knowledge.* La Salle: Open Court 1987.

[6] Riedl, R., Delpos, M. (eds.): *Die Evolutionäre Erkenntnistheorie im Spiegel der Wissenschaften*. Wien: WUV Universitätsverlag 1996.

[7] Kováč, L.: Úvod do kognitívnej biológie. *Biol. listy* **51** (1986) 172-190.

[8] Kováč, L.: The future of bioenergetics. *EBEC Reports* **4** (1986) 26-27.

[9] Kováč, L.: Overview: Bioenergetics between chemistry, genetics and physics. *Curr. Zopics Bioenerg*. **15** (1987) 331-372.

[10] Goodwin, B.: *Analytical physiology of cells and developing organisms*. London: Academic Press 1976.

[11] Kuhn, H.: Selbsorganisation molekularer Systeme und die Evolution des genetischen Apparats. *Angew. Chem*. **18** (1972) 838-862.

[12] Kuhn, H.: Origin of life and physics: Diversified macrostructure - Inducement to form information-carrying and knowledge-accumulating systems. *J. Res. Develop.* **32** (1988) 37-46.

[13] Popper, K.: *The logic of scientific discovery*. London: Hutchinson 1957. (Po prvý raz uverejnené vo Viedni v 1934 s titulom *Logik der Forschung.*)

[14] Vollmer, G.: *Evolutionäre Erkenntnistheorie*. Stuttgart: Hirzel 1987. (Prvé vydanie v 1975.)
[15] Lorenz, K.: *Die Rückseite des Spiegels*. München: Piper 1973.
[16] Riedl, R.: *Biologie der Erkenntnis*. Berlin: Parey 1979.
[17] Riedl, R.: *Wahrheit und Wahrscheinlichkeit*. Berlin: Parey 1992.
[18] Georgescu-Roegen, H.: *The entropy law and the economic process.* Cambridge, Mass.: University Press 1971.
[19] Lorenz, K.: Die angeborenen Formen möglicher Erfahrung. *Z. Tierpsychologie* **5** (1943) 235-409.
[20] Vollmer, G.: Mesokosmos und objektive Erkenntnis - Über Probleme, die von der evolutionären Erkenntnistheorie gelöst werden, pp. 29-91. In: Lorenz, K., Wuketits, F. M. (eds.): *Die Evolution des Denkens*. München: Piper 1983.
[21] Bohr, N.: *Atomphysik und menschliche Erkenntnis*, vol. I. Braunschweig: Vieweg 1964.
[22] Simon, H. A.: *Reason in human affairs*. Stanford: Stanford University Press 1983.
[23] Feynman, R. P., Leighton, R. B., Sands, M.: *The Feynman lectures in physics*. Vol. III. Reading: Addison-Wesley 1966.
[24] Eigen, M., Winkler, R. (1975) *Das Spiel*. München: Piper 1975.
[25] Schuster, P., Sigmund K.: Replicator dynamics. *J. Theor. Biol*. **100** (1983) 533-538.
[26] Lorenz, K.: Kants Lehre vom Apriorischen im Lichte gegenwertiger Biologie, pp. 95-124. In: Lorenz, K., Wuketits, F. M. (eds.): *Die Evolution des Denkens.* München: Piper 1983. Po prvý raz uverejnené v *Blätter für deutsche Philosophie* **15** (1941) 94-125.
[27] Campbell, D. T.: Evolutionary epistemology, pp. 413-463. In: Schilpp, P. A. (ed.): *The philosophy of Karl Popper*. La Salle: Open Court 1974.
[28] Riedl, R.: Deficiencies of adaptation in human reason; a constructivist extension of evolutionary epistemology. *Evol. Cogn*. **1** (1995) 27-37.
[29] Nagel, E., Newman, J. R.: *Gödel´s proof.* Paul: Routledge and Kegan 1958.
[30] Hofstadter, D. A.: *Gödel, Escher, Bach: An eternal golden braid*. Harmondsworth: Penguin Books 1980.
[31] Penrose, R.: *Shadows of the mind*. Oxford: Oxford University Press 1994.
[32] Chaitin, G. J.: *Algorithmic information theory*. Cambridge: Cambridge University Press 1990.
[33] Kolgomorov, A. N.: Three approaches to the quantitative definition of information. *Problemy peredači informciji* **1**(1) (1965) 3-11.
[34] Chaitin, G. J.: Randomness and mathematical proof. *Sci. Amer*. **232** (5) (1975) 47-52.
[35] Tarski, A.: The semantic conception of truth and the foundations of semantics, pp. 52-84. In: Feigl, H., Sellars, W. (eds.): Readings in philosophical analysis. New York: Appleton-Century-Crofts 1949.
[36] Penrose, R.: *The Emperor´s new mind*. Oxford: Oxford University Press 1986.
[37] Kuhlenbeck, H.: *The human brain and its universe*. Basel: Karger 1982.
[38] Gerlach, J.: Gödel-Theorem and Kuhlenbeck-Paradox. *Naturwiss*. **75** (1988) 393-398.
[39] Stent, G. S.: *Molecular biology of bacterial viruses*. San Francisco: Freeman 1963.
[40] *C. elegans* sequencing consortium (1998) Genome sequence of the nematode *C. elegans.* A platform for investigating biology. *Science* **282** (1998) 2012-2018.

[41] Bargmann, C. I.: Neurobiology of the *Caenorhabditis elegans* genome. *Science* **282** (1998) 2028-2033.
[42] Heschl, A.: *Das intelligente Genom*. Berlin: Springer 1998.
[43] Vollmer, G.: *Was kőnnen wir wissen?* Band 1. Die Natur der Erkenntnis. Stuttgart: Hirzel 1985.
[44] Fontana, W., Wagner, G., Buss, L. W.: Beyond digital naturalism, pp. 211-227. In: Langton, C. G. (ed.): *Artificial Life.* Cambridge, Mass.: Massachusetts Institute of Technology Press 1997.
[45] Prigogine, I.: *Introduction to thermodynamics of irreversible processes*. New York: Wiley 1967.
[46] Spiegelman, S.: An approach to the experimental analysis of precellular evolution. *Quart. Rev. Biophysics* **4** (1971) 213.
[47] Küppers, B.: Towards an experimental analysis of molecular self-organization and precellular Darwinian evolution. *Naturwiss*. **66** (1979) 228-243.
[48] Jacob, F.: *Le jeu des possibles*. Paris: Fayard 1981.
[49] Lloyd, S., Pagels, H.: Complexity as thermodynamic depth. *Ann. Phys*. **188** (1988) 186-213.
[50] Grassberger, P.: Problems in quantifying self-generated complexity. *Helv. Physica Acta* **62** (1989) 489-511.
[51] Gell-Mann, M.: *The quark and the jaguar*. New York: Freeman 1994.
[52] Riedl, R.: *Darwin, Zeus und Russells Huhn*. Wien: Kremayr & Scheriau 1994.
[53] Popper, K.: Die erkenntnistheoretische Position der evolutionären Erkenntnistheorie, pp. 29-37. In: Riedl, R., Wuketits, F. W. (eds.): *Die Evolutionäre Erkenntnistheorie.* Berlin: Parey.
[54] Drake, F. D.: *The Drake equation: a reprisal*. New York: Byron Press 1990.
[55] Gott III, J. R.: Implications of the Copernican principle for our future prospects. *Nature* **363** (1993) 315-319.
[56] Feynman, R. P., Leighton, R. B., Sands, M.: *The Feynman lectures in physics*. Vol. II. Reading: Addison-Wesley 1966.
[57] Magnasco, M. O.: Forced thermal ratchets. *Phys. Rev. Lett*. **10** (1993) 1477-1481.
[58] Peskin, C. P., Garrett M. O., Oster G. F.: Cellular motions and thermal fluctuations: the Brownian ratchet. *Biophys. J*. **65** (1993) 316-324.
[59] Astumian, R. D.: Thermodynamics and kinetics of a Brownian motor. *Science* **276** (1997) 917-922.
[60] Maynard Smith, J.: *Evolutionary genetics*. Oxford: Oxford University Press 1989.
[61] Fontana, W., Schuster, P.: Continuity in evolution: On the nature of transitions. *Science* **280** (1998) 1451-1455.
[62] Nicolis, G., Prigogine, I.: *Die Erforschung des Komplexen*. München: Piper 1987.
[63] Dawkins, R.: *The blind watchmaker*. New York: Norton 1987.
[64] Edelman, G. M.: *Neural darwinism*. New York: Basic Books 1987.
[65] Kupiec, J. J.: A Darwinian theory for the origin of cellular differentiation. *Mol. Gen. Genet*. **255** (1997) 201-208.
[66] Britten, R. J._ Underlying assumptions of developmental models. *Proc. Natl. Acad. Sci. U. S.* **95** (1998) 9372-9377.
[67] Maes, P.: Modelling adaptive autonomous agents. In: Laghton C. G. (ed.): *Artificial life*. Cambridge, Mass.: MIT Press 1997.

[68] Jacob, F.: Evolution and tinkering. *Science* **196** (1977) 1161-1167.
[69] Wimsatt, W. C.: Developmental constraints, generative entrenchment, and the innate-acquired distinction. In: Bechtel, W. (ed.): *Integrating scientific disciplines.* Dordrecht, Nijhoff 1986.
[70] Desplan, C.: Eye development: governed by a dictator or a junta? *Cell* **91** (1997) 861-864.
[71] Kováč, L.: Prirodzený svet („životný svet") ako biologický problém. *Vesmír* **71** (1992) 505-506.
[72] Kováč, L.: European culture in the global conflict of cultures: a view of a biologist. In: Fukač, J., Chlup, Z., Mizerová, A., Schauerová, A. (eds.): *The crossroads of European culture*. Brno: Vutium Press 1999.
[73] Russell, B.: *A history of western philosophy*. London: Unwin paperbacks 1961.
[74] Mach, E. (1923) Die ökonomische Natur der physikalischen Forschung, pp. 217-244. In: *Populärwissenschaftliche Vorlesungen*. Leipzig: Barth 1923.
[75] Sober, E.: Parsimony, pp. 249-254. In: Keller, E. F., Lloyd, E. A. (eds.): *Keywords in evolutionary biology*. Cambridge, Mass.: Harvard University Press 1992.
[76] Jaynes, E. T.: Information theory and statistical mechanics. *Phys. Rev*. **106** (1957) 620-630.
[77] Nagel, T.: What it is like to be a bat. *The Philosophical Review* **83** (1974) 435-451.
[78] Plutchik, R.: Emotions and evolution. *Internat. Rev. of Studies on Emotion* **1** (1991) 37-58.
[79] Damasio, A. R.: *Descartes´ error. Emotion, reason, and the human brain*. New York: Putnam's sons 1994.
[80] Wimmer, M.: Evolutionary roots of emotions. *Evol. Cogn*. **1** (1995) 38-50.
[81] Kováč, L.: Evolution of affective evaluation of external stimuli, pp. 867-874. In: Novák, V. J. A., Mlíkovský, J. (eds.): *Evolution and environment*. Prague: Czechoslovak Academy of Sciences 1982.
[82] Buber, M.: *Ich und Du*. Leipzig: Insel Verlag 1923.
[83] Pauling, L.: Scientists in politics, pp. 97-105. In: Tiselius, A., Nilsson, S. (eds.): *The place of values in a world of facts*. Stockholm: Almquist and Wiksell 1970.
[84] Planck, M.: Das Prinzip der kleinsten Wirkung, pp. 91-101. In: *Vorträge und Reden*. Braunschweig: Vieweg 1958.
[85] Landau, L. D., Lifšic, J. M.: *Kratkij kurs teoretičeskoj fiziki*. Vol. 1. Moskva: Nauka 1969.
[86] Taatgen, N.: The atomic components of thought. *Trends in Cogn. Sci*. **3** (1999) 82.

Molekulárna darwinovská evolúcia *in silico*

Vladimír Kvasnička[1]

Abstrakt. V práci je prezentovaný jednoduchý model molekulárnej Darwinovej evolúcie implementovaný pomocou metód umelej chémie. Chemický reaktor (chemostat) obsahuje molekuly, ktoré sú reprezentované binárnymi reťazcami, ktoré sú schopné replikácie s pravdepodobnosťou úmernou ich replikačnej rýchlostnej konštante (fitnes). Replikačný proces nie je bezchybný, obsahuje malé nepresnosti (mutácie), ktoré produkujú nové reťazce blízke pôvodným reťazcom, z ktorých vznikli nové reťazce procesom replikácie. Dynamické vlastnosti chemostatu sú špecifikované Eigenovým systémom diferenciálnych rovníc, ktorý má jediné asymptotické riešenie obsahujúce len replikátory s maximálnou replikačnou rýchlostnou konštantou (fitnes). Replikačná rýchlostná konštanta je špecifikovaná pomocou podobnosti medzi foldingom (zvinutím) daného reťazca a požadovaným cieľovým foldingom. Metóda poskytuje detailný pohľad na mechanizmy molekulárnej Darwinovej evolúcie.

Motto:
Tak, ako to vidím ja, centrálnym problémom evolúcie je mechanizmus, pomocou ktorého druhy môžu spojite nájsť cestu z nižších na vyššie vrcholy.
(*Sewall Wright* [36])

1 Úvod

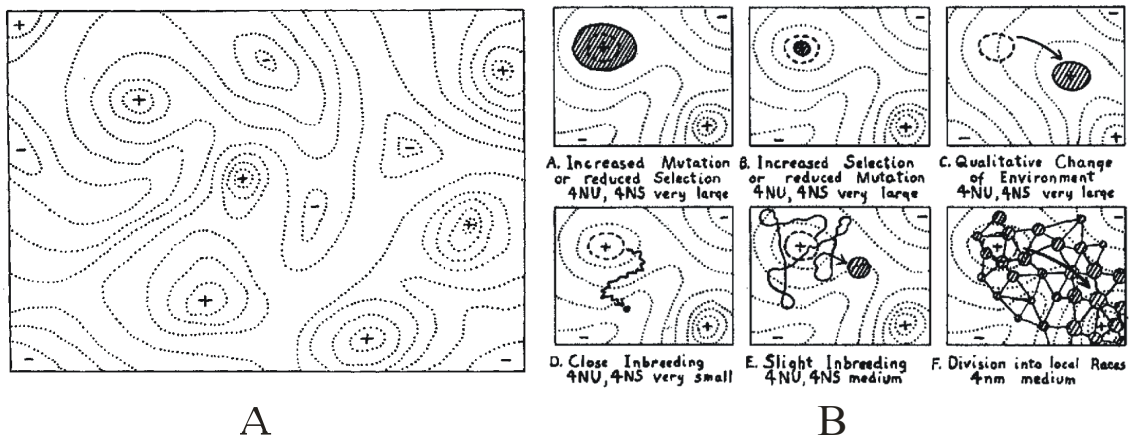
Darwinovská evolúcia patrí medzi štandardné objekty štúdia ***umelého života*** (AL), pričom hlavný stimul pre túto skutočnosť bol prudký rozvoj evolučných algoritmov na prelome 80-90. rokov minulého storočia. Táto skutočnosť má svoj počiatok u teoretického populačného genetika a evolučného biológa Sewalla Wrighta [36], ktorý v r. 1932 publikoval prácu, v ktorej zaviedol pojem „adaptive landscape“[2] a charakterizoval Darwinovu evolúciu ako adaptívny proces (v matematickej terminológii „optimalizačný proces“), kde genotyp populácie ja adaptovaný tak, aby dosiahol lokálne (snáď aj globálne) optimum na povrchu fitnes funkcie (pozri obr. 1). Omnoho neskoršie, po 40. rokoch, využil informatik John Holland [17] túto metaforu Darwinovej evolúcie pre návrh genetických algoritmov, ktoré zo súčasného pohľadu môžu byť interpretované ako abstrakcia Darwinovej evolúcie do formy univerzálneho optimalizačného algoritmu[3, 4].

---

[1] Fakulta informatiky a informačných technológií, Slovenská technická univerzita, 842 16 Bratislava,
E-mail: kvasnicka@fiit.stuba.sk

[2] V súčasnosti sa používa angl. termín „fitness landscape“. Vychádzajúc zo súčasnej slovenskej (a českej) matematickej terminológie, kde je zaužívaný termín „(hyper)povrch funkcie“ (ako preklad angl. „function (hyper)surface“), používame preklad tohto termínu ako „povrch fitnes“. Týmto sa chceme vyhnúť krkolomným prekladom termínu „landscape“.

[3] Americký filozof a kognitívny vedec Daniel Dennett vo svojej knihe *Darwin's Dangerous Idea - Evolution and the Meaning of Life* [8] univerzálnosť Darwinovej evolúcie zdôraznil tým, že ju charakterizoval ako univerzálny algoritmus.



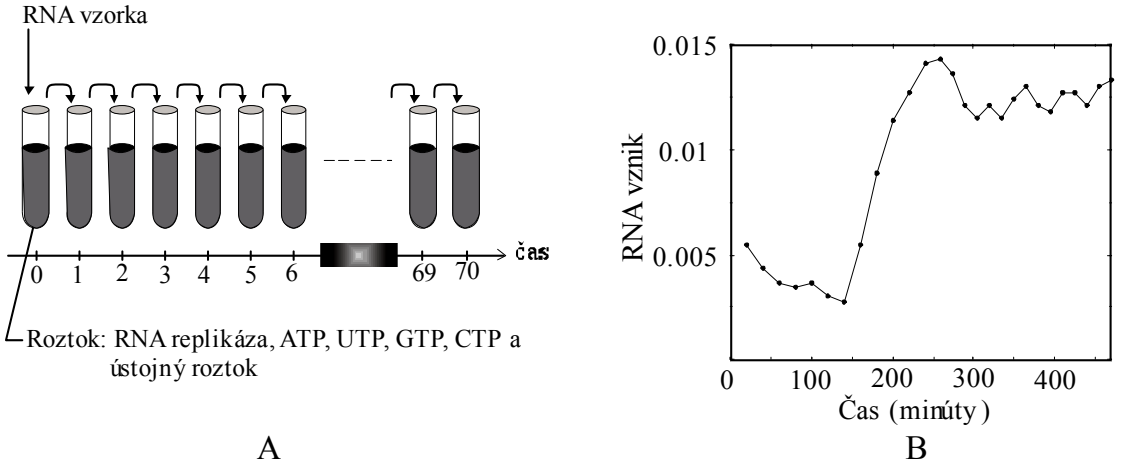
Obrázok 1. Jedny z najslávnejších obrázkov neodarwinizmu, ktoré pochádzajú od amerického populačného genetika a evolučného biológa Sewalla Wrighta [36]. (A) Znázorňuje povrch fitnes funkcie obsahujúci množstvo lokálnych extrémov (miním a maxím). Evolúciu charakterizoval ako adaptívny proces (v matematickej terminológii „optimalizačný proces“), kde genotyp populácie ja adaptovaný tak, že dosiahne lokálne (snáď aj globálne) optimum – maximum na povrchu fitnes funkcie. (B) Wright navrhoval rôzne mechanizmy, ako prekonať hlboké údolia medzi dvoma maximami. Svoju pozornosť v tomto smere sústredil hlavne na možnosť malej populácie, ktorá je dočasne izolovaná od hlavnej populácie a kde nadobúda význam genetický drift s poklesom významu prirodzeného výberu. V takomto evolučnom „taviacom kotle“ dochádza k neočakávaným zmenám genotypu, čo sa môže prejaviť aj preskokom z jedného maxima na druhé maximum. Týmto spôsobom vysvetľoval Wright mechanizmus toho (pozri motto tohto príspevku), ako sa genotyp populácie môže presunúť z nižšieho vrcholu na vyšší vrchol povrchu fitnes.

Cieľom tohto príspevku je prezentácia jednoduchého modelu Darwinovej evolúcie, ktorý postihuje jej elementárne aspekty a procesy vyskytujúce sa na úrovni biomakromolekúl. Diskutovaný model evolúcie poskytuje semikvantitatívnu teóriu, ktorá je schopná interpretovať mnohé jej pojmy a koncepty, ako je napr. úloha neutrálnych mutácií v priebehu adaptačného procesu. Dôležitú úlohu v tomto prístupe hrajú metódy ***umelej chémie*** [2,4,5,6,9,14,16,26], ktorá je v súčasnosti považovaná za oblasť úzko súvisiacu s umelým životom. Prístup umelej chémie je založený na metafore chemostatu (chemického reaktoru), ktorý obsahuje „molekuly“ reprezentované abstraktnými objektami (reťazcami znakov, grafmi, formulami z nejakého jazyka, atď.). Molekuly sú v chemostate transformované pomocou „chemických reakcií“ na iné prípustné objekty, pričom pravdepodobnosť týchto transformácií je striktne určená štruktúrou reagujúcich objektov. Kinetika procesov prebiehajúcich v chemostate je špecifikovaná Eigenovými replikátorovými diferenciálnymi rovnicami [10,11], ktoré majú svoj „zdravý“ základ vo fyzikálnej chémii (zákon účinných hmotností [1]). Hlavným predmetom záujmu umelej chémie je (*i*) štúdium formálnych systémov založených na metafore chemostatu, a ktoré sú schopné vykonávať špeciálne vysoko paralelné výpočty a (*ii*) vytvárať formálne

[4] Známy anglický evolučný biológ a popularizátor Richard Dawkins v eseji *Universal Darwinism* [7] charakterizoval Darwinovu evolúciu ako univerzálny prístup k spontánnemu vytváraniu adaptívnej zložitosti (adaptive complexity), ktorá je charakteristická pre život v celom našom vesmíre. Problém univerzálnosti Darwinovej evolúcie je u nás široko diskutovaný filozofom Dušanom Gálikom v jeho prednáške [15], v ktorej kriticky hodnotí možnosti univerzálneho darwinizmu ako východiska pre zjednotenie prírodných a spoločenských vied.

autokatalytické systémy (molekuly sú reprezentované štruktúrovanými objektmi) pre účely *in-silico* simulácií emergencie „živých“ systémov. Folding binárnych reťazcov (ktoré chápeme ako genotyp) interpretujeme ako fenotyp, pomocou ktorého je špecifikovaný fitnes molekúl. Táto jednoduchá triáda ***genotyp – fenotyp – fitnes*** má fundamentálny význam v našich počítačových simuláciách molekulárnej Darwinovej evolúcie, ktorá je chápaná ako postupná zmena fenotypu reťazcov.

Viedenský teoretický chemik a biológ Peter Schuster už viac ako 10 rokov študuje Darwinovu evolúciu na molekulárnej úrovni pomocou RNA molekúl [14,31,32], ktorých fitnes je určené ich foldingom. Hlavný význam týchto prác spočíva v tom, že poskytujú teoretický pohľad na Darwinovu evolúciu na elementárnej molekulárnej úrovni. Ich výsledky poskytujú argumentačný aparát pre vysvetlenie základných mechanizmov Darwinovej evolúcie biomakromolekúl, menovite ukazujú možné mechanizmy náhlej zmeny fenotypu (foldingu molekuly) pomocou postupnosti 1-bodových mutácií a taktiež význam neutrálnych mutácii pre evolúciu. Cieľom našich prác [21-27] v tejto oblasti bolo zopakovať výsledky Schustra a spol. pre zjednodušený model biomakromolekuly, ktorá je teraz reprezentovaná binárnym reťazcom v rámci umelej chémie.



Obrázok 2. Schematické znázornenie Spiegelmanovho experimentu [33]. (A) Skúmavky obsahujú roztok RNA replikázy (enzým, pomocou ktorého sa replikuje RNA molekula), základné stavebné jednotky RNA: adenozíntrifosfát (ADP), uraciltrifosfát (UTP), guaníntrifosfát (GTP), cytozíntrifosfát (CTP) a ústojný roztok, ktorý fixuje v skúmavke približne neutrálne *pH*. Experiment prebieha tak, že do prvej skúmavky zľava pridáme vzorku RNA, po určitom čase (niekoľko hodín) sa v skúmavke rozmnoží RNA molekula. Z prvej skúmavky vyberieme 1 ml roztoku a prenesieme ho do druhej skúmavky, po rovnakom čase ako v predošlom prípade necháme v tejto skúmavke rozmnožiť RNA, zo skúmavky odoberieme 1 ml vzorky a prenesieme ho do tretej skúmavky. Tento proces „pasážovania“ opakujeme 70-krát, pričom v každej skúmavke vyhodnocujeme množstvo vzniknutej RNA procesom replikácie. (B) Diagram znázorňuje efektivitu replikačného procesu v rôznych skúmavkách. V prvej etape bola aktivita replikačného procesu pomerne malá, v čase okolo 200 minút sa náhle zväčšila o jeden rád, a potom až do konca experimentu aktivita zostala konštantná. Táto náhla zmena aktivity replikácie RNA sa vysvetľuje tým, že pri procese replikácie RNA nastala mutácia, ktorá podstatne zvýšila aktivitu replikácie RNA molekuly.

Pred nedávnom Newman a Engelhardt [30] ukázali, že použitím Kauffmanovej *KN* funkcie[5] [19,20] môžu byť zreprodukované skoro všetky základné výsledky získané Petrom Schusterom a jeho spolupracovníkmi [14,31,32] o molekulárnej Darwinovej evolúcii, ktoré sú založené na použití „fyzikálneho“ modelu RNA a jej foldingu. Bolo ukázané, že použitie tohto jednoduchého modelu povrchu fitnes funkcie poskytuje výsledky, ktoré sú taktiež formálne veľmi blízke tým, ktoré sú predpovedané replikátorovým modelom. V našich výpočtoch tohto zamerania sme navrhli [26] jednoduchý model fitnes funkcie pre simuláciu molekulárnej Darwinovej evolúcie. Získané výsledky sú veľmi blízke teoretickým výsledkom, ktoré sú založené na Eigenovej teórii replikátorov.

2 Eigenove replikátory

Na prelome 60. a 70. rokov publikoval Manfred Eigen významnú prácu „*Self organization of matter and the evolution of biological macro molecules*“ [10], v ktorej študoval vlastnosti chemického systému zloženého z tzv. *replikátorov* (priamo v texte tejto práce uvádza, že bol stimulovaný k jej napísaniu Spiegelmanovými [33] experimentami *in-vitro* Darwinovej evolúcie, pozri taktiež ref. [30]). Replikátory sú hypotetické biomakromolekuly, ktoré majú schopnosť replikovať sa. Eigen ukázal, že takýto jednoduchý chemicko-kinetický systém formálne simuluje Darwinovu evolúciu na molekulárnej úrovni. Ďalšie zovšeobecnenie modelu replikátorov bolo vykonané Eigenom a Schustrom [11] v sérii prác so spoločným názvom „*The Hypercycles: A Principle of Natural Evolution*“, v ktorých sa položili základy teórie *hypercyklov*, ktoré sú v súčasnosti považované za jeden z nemnohých formálnych systémov schopných simulovať nárast zložitosti (komplexity) na rozhraní živých a neživých systémov.

Študujme molekuly - replikátory $X_1, X_2, ..., X_n$, ktoré sú schopné nasledujúcich dvoch chemických reakcií

$$X_i \xrightarrow{k_i} X_i + X_i \quad (i = 1, 2, ..., n) \tag{1a}$$

$$X_i \xrightarrow{\phi} \varnothing \quad (i = 1, 2, ..., n) \tag{1b}$$

Prvá reakcia (1a) znamená, že molekula X_i sa replikuje s rýchlostnou konštantou k_i a druhá reakcia (1b) znamená, že molekula X_i zaniká s rýchlostnou konštantou ϕ (ktorá sa nazýva *zrieďovací tok* a bude špecifikovaná neskoršie). Ak použijeme k týmto reakciám štandardný fyzikálno-chemický zákon účinných hmotností [1], dynamika kinetického systému obsahujúceho reakcie (1a-b) je určená diferenciálnymi rovnicami

$$\dot{x}_i = x_i \left(k_i - \phi \right) \quad \left(i = 1, 2, ..., n \right) \tag{2a}$$

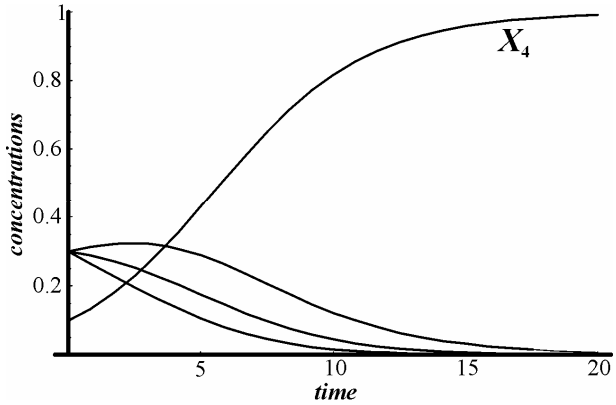
kde $0 \le x_i \le 1$ je koncentrácia molekuly X_i. Zrieďovací tok ϕ je chápaný ako „voľný parameter“, ktorý sa nastaví tak, aby suma časových derivácií koncentrácií bola nulová, $\sum \dot{x}_i = 0$

[5] Táto stochastická funkcia bola vytvorená [19,20] teoretickým biológom Stuartom Kauffmanom pomocou fyzikálnej „metafory“ spinových skiel. V súčasnosti je veľmi populárna v AL a teoretickej biológii, ako jednoduchý prostriedok pre numerickú implementáciu „rozoklaných“ (rugged) povrchov fitnes, ktorých členitosť sa dá jednoducho „nastavovať“ pomocou niekoľkých celočíselných parametrov.

$$\dot{x}_i = x_i\left(k_i - \sum_{j=1}^{n} k_j x_j\right) \quad (i = 1,2,...,n) \tag{2b}$$

kde podmienka $\sum x_i = 1$ bola použitá bez straty všeobecnosti našich úvah. Analytické riešenie systému diferenciálnych rovníc má tvar [10]

$$x_i(t) = \frac{x_i(0)e^{k_i t}}{\sum_{j=1}^{n} x_j(0)e^{k_j t}} \tag{3}$$



Obrázok 3. Časový priebeh koncentrácií 4-zložkového systému s rýchlostnými konštantami k_1=1, k_2=2, k_3=3 a k_4=4. Molekuly X_4 víťazia, aj keď ich počiatočná koncentrácia bola najmenšia, v čase t>15 systém už obsahuje skoro výlučne len molekuly X_4, ostatné vymizli.

Toto riešenie vyhovuje asymptotickej podmienke, kde „prežíva“ len tá molekula, ktorá má maximálnu rýchlostnú konštantu k_{max}, zatiaľ čo ostatné molekuly vymiznú (pozri obrázok 3)

$$\lim_{t\to\infty} x_i(t) = \begin{cases} 1 & (\text{pre } k_i = k_{max} = max\{k_1,...,k_n\}) \\ 0 & (\text{ostatné prípady}) \end{cases} \tag{4}$$

Každý typ molekuly je formálne chápaný ako zvláštny biologický druh s fitnes špecifikovanou rýchlostnou konštantou k. Potom prežívajú len tie molekuly - biologické druhy, ktoré sú najlepšie prispôsobené, t.j. majú najväčšiu rýchlostnú konštantu k_{max}; všetky ostatné molekuly - druhy s rýchlostnými konštantami fitnes menšími ako k_{max} vymiznú v priebehu časovej evolúcie chemicko-kinetického systému, pozri obr. 3. Podmienka konštantnosti sumy koncentrácie (t.j. $\sum x_i = 1$) zavádza do systému dodatočný „selekčný tlak“, replikujúce sa molekuly vytesňujú ostatné molekuly.

Eigenov model replikátorov môže byť jednoducho modifikovaný tak, že sa zavedú mutácie do procesu replikácie (1a)

$$X_j \xrightarrow{k_{ji}} X_j + X_i \quad (i, j = 1,2,...,n) \tag{5}$$

kde k_{ji} je rýchlostná konštanta tejto reakcie. Predpokladá sa, že matica rýchlostných konštánt $\boldsymbol{K} = (k_{ij})$ má dominantné diagonálne elementy, t.j. nediagonálne elementy sú podstatne (rádovo) menšie ako diagonálne elementy ($k_{ij} \ll k_{ii}$, pre i,j=1,2,...,n a $i \neq j$).

Tento dôležitý predpoklad vyplýva priamo zo skutočnosti, že „nepresné“ replikácie (5) pre $i \neq j$ sú veľmi vzácne, t.j. produkt X_j je uvažovaný ako mutácia eduktu X_i, $X_j = O_{mut}\left(X_i\right)$. Systém diferenciálnych rovníc (2a) má potom tvar

$$\dot{x}_i = x_i\left(k_{ii} - \phi\right) + \sum_{j \neq i} k_{ji} x_j \quad \left(i = 1, 2, ..., n\right) \tag{6}$$

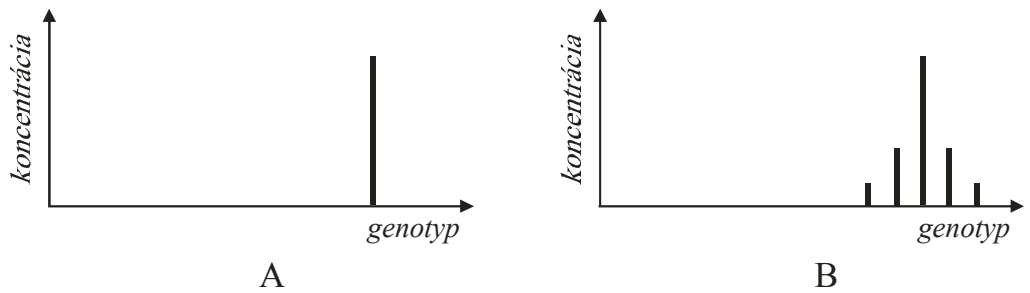
Zrieďovací tok ϕ z (6) je určený tak, aby aj v tomto všeobecnejšom prípade suma časových derivácií koncentrácií bola nulová ($\sum \dot{x}_i = 0$, alebo $\sum x_i = 1$)

$$\phi = \sum_{i,j=1}^{n} k_{ij} x_j \tag{7}$$

Analytické riešenie (6) má tvar [18]

$$x_i\left(t\right) = \frac{\sum_{j=1}^{n} q_{ij} x_i\left(0\right) e^{\lambda_i t}}{\sum_{m,j=1}^{n} q_{mj} x_j\left(0\right) e^{\lambda_j t}} \tag{8}$$

kde $\boldsymbol{Q} = \left(q_{ij}\right)$ je nesingulárna matica, ktorá diagonalizuje maticu rýchlostných konštánt $\boldsymbol{K}$, $\boldsymbol{Q}^{-1}\boldsymbol{K}\boldsymbol{Q} = \boldsymbol{\Lambda} = diag\left(\lambda_1, \lambda_2, ..., \lambda_n\right)$. Pretože bolo postulované, že matica $\boldsymbol{K}$ má dominantné diagonálne elementy, jej vlastné hodnoty λ sú blízke diagonálnym elementom, $\lambda_i \doteq k_{ii}$, a transformačná matica $\boldsymbol{Q}$ je blízka jednotkovej matici, $q_{ij} \doteq \delta_{ij}$ (Kroneckerov delta symbol). To znamená, že zavedenie slabých mutácií nemení dramaticky všeobecné vlastnosti pôvodného jednoduchého systému replikátorov (1-4) bez mutácií. Menovite, v asymptotickom čase $t \to \infty$ systém obsahuje skoro výlučne len molekuly majúce najväčšie rýchlostné konštanty k_{max}, ktoré sú vzácne doprevádzané inými replikátormi tvorenými z víťazného replikátora (s rýchlostnou konštantou k_{max}) pomocou nepresného replikačného procesu (5), pozri obr. 4.



Obrázok 4. (A) Asymptotickú situáciu pre $t \to \infty$ pre systém replikátorov popísaných reakciami (1a-b), t.j. ktoré sa presne replikujú. V tomto prípade asymptoticky stabilný stav systému obsahuje len replikátory s maximálnou rýchlostnou konštantou k_{max}. (B) Asymptoticky stabilný stav systému, kde replikácia nie je už presný proces, kde sa už uvažujú mutácie pri replikačnom procese. Tento stabilný stav je zložený z niekoľkých replikátorov, pričom najväčšia koncentrácia prislúcha replikátoru s maximálnou rýchlostnou konštantou k_{max}, ostatné replikátory sa vyskytujú s podstatne menšou koncentráciou. Replikátory vyskytujúce sa súčasne s hlavým víťazom X_{max} sú tie, ktoré sú z neho tvorené mutáciami.

3 Metafora chemického reaktoru - chemostat

Uvažujme chemický reaktor - *chemostat*, ktorý obsahuje formálne objekty nazývané molekuly, ktoré sú reprezentované reťazcami znakov. Chemostat nie je priestorovo štruktúrovaný, v chémii sa hovorí, že je s dobrým miešaním (well stirred reactor), pozri obr. 5. Akt chemickej reakcie - replikácie spočíva v nasledujúcich troch krokoch:

(1) náhodný výber replikátora M (ktorý je totožný s X_i) z chemostatu,

(2) pre takto vybraný replikátor sa uskutoční s pravdepodobnosťou úmernou rýchlostnej konštante k_{ii} proces replikácie, pričom vznik chybného replikátora M' (ktorý je totožný s X_j) je uskutočnený s pravdepodobnosťou úmernou rýchlostnej konštante k_{ij}.

(3) Vzniknutým novým replikátorom M' (bez mutácie alebo s mutáciou) sa nahradí iný náhodne vybraný replikátor z chemostatu (t.j. vo všetkých elementárnych krokoch máme v chemostate konštantný počet replikátorov).

Počítačová simulácia chemostatu a v ňom prebiehajúce replikačné reakcie sa dá pomerne jednoducho implementovať pomocou jednoduchého stochastického prístupu [16]. Tento prístup je založený na skutočnosti, že ak náhodne vyberieme z chemostatu nejaký replikátor, potom pravdepodobnosť toho, že tento replikátor je totožný s nejakým daným replikátorom X, sa jednoducho rovná koncentrácii tohto replikátora v chemostate, $0 \le c(X) \le 1$. Pravdepodobnosť uskutočnenia replikácie sa rieši tak, že táto pravdepodobnosť je úmerná replikačnej rýchlostnej konštante. Algoritmicky môžeme chemostat implementovať pomocou pseudo-Pascalu týmto jednoduchým algoritmom:

Algoritmus 1

```
chemostat:=a multiset composed of randomly
           generated replicators X;
t:=0;
while t<t_max then
begin t:=t+1;
      X:=O_select(chemostat);  {selection}
      if random<prob_rep(X) then
      begin X':=X;            {replication}
            X'':=O_mut(X');   {mutation}
            X''':=O_select(chemostat);
            chemostat:=(chemostat+{X''})-{X'''};
      end;
end;
```

Algoritmus je inicializovaný náhodným vygenerovaním chemostatu. Pravdepodobnosť $\mathtt{prob}_{\mathtt{rep}}(\mathtt{X})$ je určená tak, že je úmerná rýchlostnej konštante replikácie X, pravdepodobnosť mutácie je zahrnutá priamo v operátore $\mathtt{O}_{\mathtt{mut}}(\mathtt{X'})$.

Všeobecná idea chemostatu bude ilustrovaná jeho použitím ako optimalizátora binárnej funkcie

$$f : \{0,1\}^N \to [0,1] \tag{9}$$

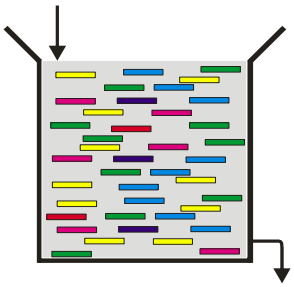
Táto funkcia mapuje binárne reťazce $\boldsymbol{x} = (x_1, x_2, ..., x_N) \in \{0,1\}^N$ dĺžky N na reálne čísla z intervalu $\langle 0,1 \rangle$. Našim cieľom bude nájsť optimálne riešenie

$$\boldsymbol{x}_{opt} = arg \max_{\boldsymbol{x} \in \{0,1\}^N} f(\boldsymbol{x}) \tag{10}$$

Pretože kardinalita množiny $\{0,1\}^N$ sa rovná 2^N, CPU čas nutný pre riešenie tejto úlohy bude rásť exponenciálne

$$t_{CPU} \approx 2^N \tag{11}$$

To znamená, že optimalizačný problém (10) patrí do triedy numericky obtiažnych NP-úplných problémov. Toto je jeden z hlavných dôvodov prečo optimalizačné problémy typu (10) sú riešené pomocou evolučných algoritmov [17,28], ktoré reprezentujú veľmi efektívne numerické techniky schopné rýchlo získať dobré suboptimálne riešenie binárnych optimalizačných problémov.



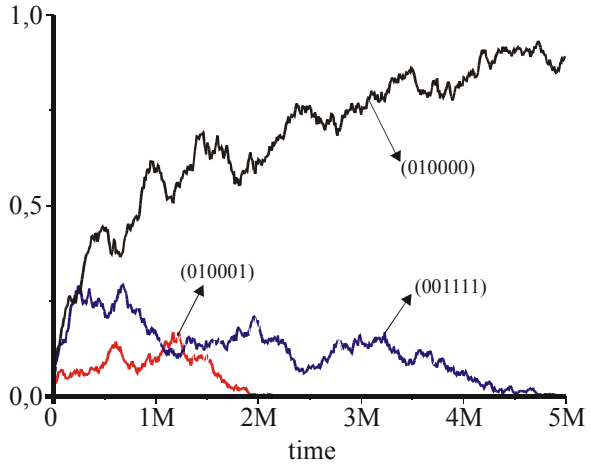
Obrázok 5. Chemostat obsahuje „zmes“ replikátorov. Replikačný akt spočíva v tom, že náhodne vybraný replikátor sa zreplikuje (buď perfektne alebo s malou mutáciou), vzniknutý produkt - replikátor - vytesní iný, náhodne vybraný replikátor. To znamená, že chemostat obsahuje v celej svojej histórii konštantný počet replikátorov.

Nech chemostat obsahuje molekuly, ktoré sú realizované binárnymi reťazcami $\boldsymbol{x} = (x_1, x_2, ..., x_N) \in \{0,1\}^N$. Uvažujme nasledujúcu monomolekulárnu reakciu - replikáciu (pozri rov. (5))

$$\boldsymbol{x} \xrightarrow{f(\boldsymbol{x})} \boldsymbol{x} + \boldsymbol{x}' \tag{12}$$

kde molekula - reaktant $\boldsymbol{x}'$ nahradí náhodne vybranú molekulu z chemostatu (pozri obrázok 5). Funkčná hodnota $f(\boldsymbol{x})$ priradená chemickej reakcii je interpretovaná ako pravdepodobnosť (rýchlostná konštanta) reakcie (12). V evolučných algoritmoch selekčný tlak v populácii riešení (chromozómov) je vytvorený pomocou fitnes. Chromozómy s väčším fitnes majú väčšiu šancu zúčastňovať sa reprodukčného procesu (miera kvality chromozómov) a naopak, chromozómy s malým fitnes len vzácne vstupujú do reprodukčného procesu. Tento jednoduchý prejav princípu Darwinovho prirodzeného výberu zabezpečuje postupnú monotónne neklesajúcu gradáciu fitnes v priebehu evolúcie populácie. V tomto prístupe zmienený princíp výberu na základe fitnes je rozšírený o dodatočný selekčný tlak založený na skutočnosti, že populácia v chemostate má konštantný počet jedincov. Molekula vstupujúca do reakcie je náhodne vybraná z chemostatu, kvázináhodne sa rozhodne, či vstupuje do reakcie alebo nie (s pravdepodobnosťou úmernou jej funkčnej hodnote), vzniknutá molekula vytesní z chemostatu inú, náhodne vybranú molekulu. Produkt $\boldsymbol{x}'$ z pravej strany reakcie (12) je

mutácia [17] vstupujúcej molekuly $\boldsymbol{x}$, $\boldsymbol{x}' = O_{mut}(\boldsymbol{x})$, kde O_{mut} je stochastický operátor mutácie, ktorý zmení jednotlivý bit s pravdepodobnosťou P_{mut}. Pseudo-Pascalovský kód tohto optimalizačného algoritmu dostaneme jednoduchou modifikáciou Algoritmu 1.



Obrázok 6. Priebeh frekvencií výskytu dominantných binárnych reťazcov dĺžky 6. Chemostat bol inicializovaný 1000 náhodne generovanými binárnymi reťazcami, po 5×10^6 krokoch dominantným riešením je optimálny reťazec $\boldsymbol{x}_{opt}$=(010000). V chemostate dlhú dobu prežíva ešte iné riešenie $\boldsymbol{x}$=(001111), real($\boldsymbol{x}$)=17/63, ktoré je susedné s optimálnym riešením, ale ktoré má od neho veľkú Hammingovu vzdialenosť (v teórii GA sa táto skutočnosť sa nazýva Hammingov).

Ako ilustračný príklad študujme chemostat pre unimodálnu funkciu určenú nad binárnymi reťazcami dĺžky 6, chemostat je multimnožina

$$P = \{\ldots,(110011),\ldots\} \subset \{0,1\}^6 \tag{13}$$

Každý binárny vektor $\boldsymbol{x}$ je ohodnotený racionálnym číslom z intervalu $\langle 0,1 \rangle$

$$real(\boldsymbol{x}) = \frac{1}{2^6-1} int(\boldsymbol{x}) \tag{14}$$

kde $int(\boldsymbol{x})$ je nezáporné celé číslo vyjadrujúce dekadickú interpretáciu binárneho reťazca $\boldsymbol{x}$. Rýchlostná konštanta k priradená reťazcu

$$k(\alpha) = f(real(\alpha)) = \frac{1}{2}\left(1 + sin\left(2\pi \cdot real(\alpha)\right)\right) \tag{15}$$

s optimálnym riešením α_{opt}=(010000), kde $real(\boldsymbol{x})$=16/63 a f(16/63)=0.999845. Chemostat je inicializovaný 1000 náhodne generovanými binárnymi reťazcami a mutačný operátor O_{mut} je špecifikovaný 1-bit pravdepodobnosťou P_{mut}=0.01. Z numerických výsledkov znázornených na obr. 6 vyplýva, že v chemostate spontánne emerguje binárny reťazec, ktorý je optimálnym riešením s racionálnou hodnotou blízkou k presnému riešeniu x_{opt}=0.25. Hlavné výsledky tejto sekcie môžu byt zhrnuté takto:

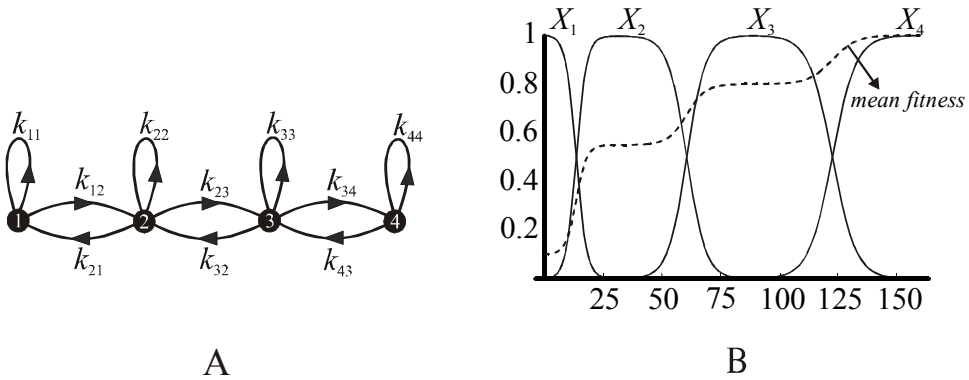
(1) Metafora Eigenových replikátorov poskytuje efektívny stochastický tzv. ***replikátorový optimalizačný algoritmus***, kde

(2) ***dôkaz konvergencieu*** tohto algoritmu ku globálnemu riešeniu bezprostredne vyplýva z existencie jediného asymptoticky stabilného riešenia s najväčšou rýchlostnou konštantou.

(3) Tento algoritmus je veľmi ***podobný*** štandardnému genetickému algoritmu [12,17,28], avšak je založený na úplne odlišnej metafore ako GA, metafora darwinovskej evolúcie v GA je nahradená metaforou chemostatu replikátorov.

4 Replikátory ako simulátory Darwinovej evolúcie

Eigenov systém replikátorov s mutáciou (t.j. s nepresnou replikáciou) študovaný v predchádzajúcej sekcii bude použitý k štúdiu molekulárnej Darwinovej evolúcie. Uvažujme hypotetický systém obsahujúci 4 replikátory X_1, X_2, X_3, a X_4. Tieto replikátory majú tú vlastnosť, že ľubovoľný z nich X_i produkuje pomocou nepresnej replikácie susedné replikátory $X_{i\pm1}$, pozri obr. 7, diagram A. Ak počiatočná koncentrácia replikátora X_1 je $x_1(0)=1$, potom v priebehu časovej evolúcie systému vznikajú koncentračné vlny, ktoré sú postupne priradené replikátorom X_2, X_3, a X_4, pozri obr. 7. diagram B.



Obrázok 7. Diagram (A) reprezentuje systém 4 replikátorov, kde replikátor X_i produkuje pomocou nepresných replikácií (mutácií) susedné replikátory X_{i-1} a X_{i+1}. Hrany grafu sú ohodnotené rýchlostnými konštantami, ktorých numerické hodnoty sú určené maticou $\boldsymbol{K}$ (18). Diagonálne rýchlostné konštanty sú medzi sebou dobre odseparované a sú omnoho väčšie ako nediagonálne rýchlostné konštanty. Diagram (B) znázorňuje priebehy koncentrácií replikátorov, ktoré tvoria postupnosť „koncentračných vĺn“. Tento diagram tiež obsahuje priebeh priemerného fitnes systému špecifikovaného výrazom $\bar{k} = k_{11}x_1 + ... + k_{44}x_4$, ktorý tvorí typickú neklesajúcu „schodovú“ funkciu.

Táto skutočnosť môže byť jednoducho interpretovaná ako prejav molekulárnej Darwinovej evolúcie, kde fitnes jednotlivých „druhov“ je určené diagonálnymi rýchlostnými konštantami k_{ii}. Evolučný proces je inicializovaný populáciou, ktorá obsahuje len replikátor X_1. V dôsledku skutočnosti, že v priebehu jeho replikácie môžu vznikať nepresné kópie - mutanty (rýchlosť ich vzniku je špecifikovaná rýchlostnou konštantou k_{12}), nasledujúci replikátor X_2 s väčším fitnes (rýchlostnou konštantou k_{22}) než akú mal jeho predchodca X_1 ($k_{11}<k_{22}$) v ďalšej etape evolúcie zvíťazí (t.j. jeho koncentrácia narastie na skoro jednotkovú hodnotu, zatiaľ čo koncentrácia X_1 klesne na veľmi malú hodnotu). Tento proces sa opakuje pre replikátor X_2, ktorý je teraz považovaný za počiatočný replikátor (hrá rovnakú úlohu ako replikátor X_1

v predchádzajúcej etape evolúcie). Pretože replikácia X_2 nie je bezchybná, s malou pravdepodobnosťou je produkovaný replikátor X_3 s väčším fitnes než aký mal jeho predchodca X_2 ($k_{33}>k_{22}$), t.j. nový „druh“ X_3 dočasne zvíťazí. Tento proces je ukončený vtedy, keď sa posledný replikátor X_4 začne vyskytovať ako dôsledok nepresnej replikácie X_3, jeho koncentrácia postupne narastie na skoro jednotkovú hodnotu, a potom už zostáva časovo nemenná (asymptoticky stabilný stav).

Tieto úvahy môžeme sformalizovať na semikvantitatívnej úrovni. Predpokladajme, že replikátorový systém v čase t_0 je v takom prechodovom stave, kde koncentrácia X_i je skoro jednotková, zatiaľ čo koncentrácie susedných replikátorov X_{i-1} a X_{i+1} sú zanedbateľne malé, t.j. $x_i(t_0)=1-2\delta$, $x_{i-1}(t_0) = x_{i+1}(t_0)=\delta$, a $x_j(t_0) = 0$ pre zostávajúce koncentrácie, pričom δ je malé kladné číslo. Zrieďovací tok $\phi(t)$ (7) je určený $\phi(t_0)=\delta(k_{11}+k_{12})+(1-2\delta)(k_{22}+k_{21}+k_{23})+\delta(k_{33}+k_{32}+k_{34})$. Potom diferenciálna rovnica (5) pre i=2 má tvar

$$\dot{x}_2 = x_2(k_{22}-\phi)+x_1k_{12}+x_3k_{32} \tag{16}$$

Ak do tejto diferenciálnej rovnice dosadíme koncentrácie v čase t_0 a použijeme predpoklad, že $\dot{x}_2(t_0)=0$ (koncentrácia replikátora X_2 v čase t_0 dosiahla maximum), dostaneme $0=-(k_{21}+k_{23})+\delta(k_{12}+k_{32})$, alebo

$$\delta=\frac{k_{21}+k_{23}}{k_{12}+k_{32}}\to 0 \quad\Rightarrow\quad k_{21}+k_{23}\ll k_{12}+k_{32} \tag{17}$$

Z predpokladu dobrej separovateľnosti koncentračných vĺn vyplývajú silné podmienky pre rýchlostné konštanty replikátorov, ktoré sú susedné s replikátorom X_2; môžeme povedať, že suma „vychádzajúcich“ rýchlostných konštánt musí byť podstatne menšia než ako suma „prichádzajúcich“ rýchlostných konštánt. Inými slovami povedané, ***„pravdepodobnosť“ vzniku daného replikátora v dôsledku nepresných autoreplikácií susedných replikátorov musí byť omnoho väčšia než ako „pravdepodobnosť“ jeho zániku v dôsledku jeho nepresnej autoreplikácie.***

Tieto semikvantitatívne úvahy boli numericky verifikované pre 4-replikátorový systém s maticou rýchlostných konštánt

$$\boldsymbol{K}=\begin{pmatrix} 0.1 & 10^{-3} & 0 & 0 \\ 10^{-7} & 0.55 & 10^{-7} & 0 \\ 0 & 10^{-11} & 0.8 & 10^{-11} \\ 0 & 0 & 10^{-7} & 1.0 \end{pmatrix} \tag{18}$$

Táto matica vyhovuje podmienkam: (*i*) Jej diagonálne maticové elementy sú omnoho väčšie než ako nediagonálne elementy, a (*ii*) nediagonálne rýchlostné konštanty vyhovujú nerovnostiam (17). Z predchádzajúcej diskusie vyplýva, že v tomto prípade môžeme očakávať darwinovské správanie sa systému. Toto očakávanie pekne súhlasí s numerickými výsledkami z obr. 6, diagram B, kde sú ukázané koncentračné profily, ktoré sú typické pre darwinovskú evolúciu. Môžeme teda na záver tejto sekcie konštatovať, že ***Eigenova fenomenologická teória replikátorov tvorí vhodný teoretický rámec pre počítačové simulačné štúdia molekulárnej Darwinovej evolúcie*** (t.j. biomakromolekúl, ktoré sú schopné replikačného procesu, akými sú RNA alebo DNA).

5 Folding binárnych reťazcov

Podobne ako pre RNA molekuly, ktoré sú schopné vytvárať 2D sekundárne štruktúry – foldingy [31], budeme študovať podobné vlastnosti aj pre binárne reťazce (pozri obrázok 7). Folding binárnych reťazcov môže byť špecifikovaný pomocou zoznamu spriahnutých dvojíc *i-j* (pre *i*<*j*) a nespriahnutých elementov *k*

$$fold\left(\boldsymbol{g}\right)=\left\{i_1-j_1,i_2-j_2,...,i_r-j_r;k_1,k_2,...,k_q\right\} \tag{19a}$$

kde dvojice sú ohraničené týmito troma podmienkami:

(1) Pre každú dvojicu *i-j* platí

$$j-i\geq 2 \tag{19b}$$

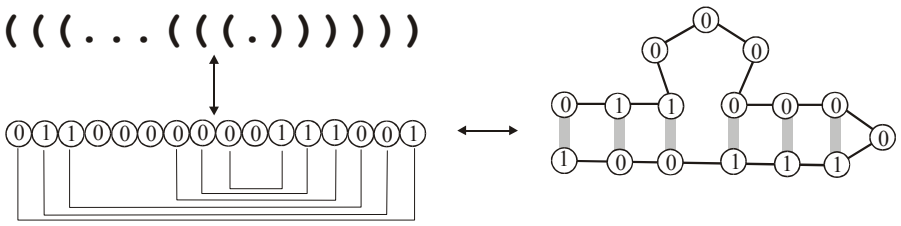
(2) Pre každé dve dvojice *i-j* a *k-l* (s podmienkou *i*≤*k*) platí buď

$$i=k \Leftrightarrow j=l \tag{19c}$$

alebo

$$k<j \Rightarrow i<k<l<j \tag{19d}$$

Prvá podmienka (19b) znamená, že minimálna dĺžka tzv. „vlasovej sponky" (po angl. *hairpin*) vo vytvorenom foldingu sa rovná dvom (pozri folding na obrázku 7). Druhá podmienka (19c) znamená, že každý element reťazca sa podieľa na vzniku maximálne jednej spriahnutej dvojice. Konečne, tretia podmienka (19d) implikuje, že tzv. „pseudouzly" sú zakázané (t. j. každý folding môže byť reprezentovaný planárnym grafom; výskyt pseudouzlov vo foldingu môže zapríčiniť ich neplanárnosť). Folding reťazca ***g*** je definovaný tak, že obsahuje maximálny počet spriahnutých párov; táto podmienka má fyzikálny význam v tom, že folding je reprezentovaný „stabilnou" sekundárnou štruktúrou. Posledné dve podmienky (19c-d) sú veľmi dôležité pre aplikáciu dynamického programovania pre konštrukciu foldingu.



Obrázok 8. Binárny reťazec (lineárny graf s vrcholmi, ktoré sú priradené binárnym elementom, ktoré sú prepojené hranami) môže byť „poskladaný" na dvojrozmernú štruktúru (ktorá sa nazýva sekundárna štruktúra) takým spôsobom, že komplementárne binárne elementy sú navzájom spriahnuté. Zátvorková reprezentácia foldingu je prezentovaná v hornej časti obrázku, kde bodové symboly sú priradené neinteragujúcim binárnym elementom a zátvorky '(' a ')' reprezentujú dvojice spriahnutých komplementárnych elementov.

Folding *fold*(***g***) binárneho reťazca ***g*** môže byť alternatívne vyjadrený pomocou zátvorkového formalizmu [27] (pozri obrázok 8). To znamená, že každý folding binárneho reťazca môže byť vyjadrený pomocou reťazca obsahujúcim symboly { **(**,**)** , **.** }; formálne *fold*(***g***)∈{ **(**,**)** , **.** }^{n}. Aktuálna forma tohto reťazca je silne ohraničená podmienkami (19b-d). Najjednoduchší spôsob ako zabezpečiť korektnosť zátvorkovej reprezentácie je použitie bezkontextovej gramatiky $\left\langle V_N,V_T,S,R\right\rangle$, kde $V_N=\left\{S\right\}$ je množina neterminálnych

symbolov, V_T={ **(**,**)**,**.**} je množina terminálnych symbolov, S je počiatočný neterminálny symbol a R je množina pravidiel, ktorá obsahuje tieto tri pravidlá

$$S \rightarrow \mathbf{(}S\mathbf{)} \mid SS \mid \mathbf{.} \tag{20}$$

Prípustný folding *fold*(***g***), ktorý spĺňa tri podmienky (19b-d) je interpretovaný ako formula obsahujúca len terminálne symboly, pričom tieto formuly tvoria jazyk L generovaný gramatikou (20), t.j. *fold*(***g***)$\in L$. Jednoduchý ilustratívny príklad produkcie formuly vyzerá takto:

$$\begin{aligned} S &\rightarrow \underline{S}S \rightarrow \mathbf{(}\underline{S}\mathbf{)}S \rightarrow \mathbf{(\,(}\underline{S}\mathbf{)\,)}S \rightarrow \mathbf{(\,(}\underline{S}\,S\mathbf{)\,)}S \rightarrow \mathbf{(\,(.}\underline{S}\mathbf{)\,)}S \\ &\rightarrow \mathbf{(\,(..)\,)}\underline{S} \rightarrow \mathbf{(\,(..)\,)}S\underline{S} \rightarrow \mathbf{(\,(..)\,)}S\underline{S}S \rightarrow \mathbf{(\,(..)\,)}S\mathbf{(}\underline{S}\mathbf{)}S \\ &\rightarrow \mathbf{(\,(..)\,)}S\mathbf{(}\underline{S}S\mathbf{)}S \rightarrow \mathbf{(\,(..)\,)}S\mathbf{(.}\underline{S}\mathbf{)}S \rightarrow \mathbf{(\,(..)\,)}\underline{S}\mathbf{(..)}S \\ &\rightarrow \mathbf{(\,(..)\,)\,.\,(..)}\underline{S} \rightarrow \mathbf{(\,(..)\,)\,.\,(..)\,.} \end{aligned}$$

Použijeme jednoduchú techniku dynamického programovania [27] k vytváraniu foldingov binárnych reťazcov. Vo všeobecnosti, tento algoritmus je alternatíva k spätnému prehľadávaniu (ktorého časová zložitosť rastie exponenciálne s dimenziou problému), vtedy, ak daný systém môže byť rozložený na také menšie podsystémy, že ohodnotenie celkového systému je súčet ohodnotení podsystémov (potom exponenciálna časová zložitosť e^n klesne na kubickú zložitosť n^3). Nech S_{ij} je podreťazec v zátvorkovej reprezentácii priradený foldingu medzi i a j (vrátane) binárnych elementov. Tieto veličiny sú inicializované takto:

$$S_{ii} = '\mathbf{.}' \quad \text{a} \quad S_{i,i+1} = '\mathbf{..}' \tag{21}$$

Nasledujúca veličina S_{ij}, pre j-i≥2, je rekurentne určená vzťahom

$$S_{i,j} = \max_{i \le k \le j-1} S_{ik} \oplus S_{k+1,j} \tag{22}$$

kde $\oplus$ je symbol spojenia „podreťazcov“ S_{ik} a $S_{k+1,j}$, ktoré už boli zostrojené v predchádzajúcej etape algoritmu (pozri Algoritmus 2). Ak S_{ik} ($S_{k+1,j}$) obsahuje v ľavej (pravej) pozícii symbol '**.**' a i-ty a j-ty binárny element sú navzájom komplementárne, potom symbol '**.**' je nahradený symbolmi **(** a **)**. Symbol „max“ v (22) znamená, že vyberieme taký index k, ktorý produkuje maximálne spriahnutie v S_{ij}. Výsledný folding je uložený v premennej S_{1n}.

Algoritmus 2

```
for i:=1 to n S[i,i]:='.';
for i:=1 to n-1 do S[i,i+1]:='..';
for d:=2 to n-1 do
for i:=1 to n-d do
begin j:=d+i;
      S[i,j]:=max{i≤k≤j-1,S[i,k]⊕S[k+1,j]};
end;
folding:=S[1,n];
```

Hlavný problém s použitím tohto algoritmu pre konštrukciu foldingu s maximálnym počtom párovaní je, že poskytuje len jeden folding z týchto maximálnych, a to ten, ktorý sa vyskytne v jeho priebehu ako prvý. Z tohto dôvodu algoritmus musí byť chápaný

ako „embryogenéza“, ktorá špecifickým spôsobom transformuje genotyp na fenotyp, čiže ako integrálna časť ohodnotenia fenotypu fitnes.

Tabuľka 1. Všetky možné genotypy, fenotypy a fitnes binárnych reťazcov dĺžky n=7

No.	Genotyp	Fenotyp	Fitnes[a]
1	(0000000) (1111111)	.....	1
2	(0000001) (0000100) (1111011) (1111110)	....(.)	1
3	(0000010) (1111101)	...(.).	0
4	(0000011) (0001001) (0001101) (0010010) (0010110) (0011000) (0011100) (0100011) (0100111) (0101001) (0110010) (0110110) (0111000) (0111100) (1000011) (1000111) (1001001) (1001101) (1010110) (1011000) (1011100) (1100011) (1100111) (1101001) (1101101) (1110010) (1110110) (1111100)	..((.))	0
5	(0000101) (0010001) (0101100) (0101110) (0110000) (0111010) (1000101) (1001111) (1010001) (1010011) (1101110) (1111010)	.((.).)	3
6	(0000110) (0011010) (1100101) (1111001)	.((.)).	5
7	(0000111) (0001111) (0010011) (0011011) (0100101) (0101101) (0110001) (0111001) (1000110) (1001110) (1010010) (1011010) (1100100) (1101100) (1110000) (1111000)	(((.)))	7
8	(0001000) (1110111)	...(..)	0
9	(0001010) (1110101)	((.).).	0
10	(0001011) (0001100) (0001110) (0011001) (0011110) (0100001) (0100100) (0100110) (0110011) (0110100) (1001011) (1001100) (1011001) (1011011) (1011110) (1100001) (1100110) (1110001) (1110011) (1110100)	.(.)(.)	0
11	(0010000) (1101111)	..(...)	1
12	(0010100) (0111011) (0111110) (1000001) (1000100) (1101011)	(.).(.)	1
13	(0010101) (0101010) (1010101) (1101010)	.((..))	1
14	(0010111) (0011101) (0100010) (0101000) (1010111) (1011101) (1100010) (1101000)	.(.(.))	0
15	(0011111) (1100000)	((.)..)	2
16	(0100000) (1011111)	.(....)	1
17	(0101011) (1010100)	(..)(.)	0
18	(0101111) (1010000)	(.(.).)	5
19	(0110101) (0110111) (1001000) (1001010)	(.)(..)	0
20	(0111101) (1000010)	(..(.))	2
21	(0111111) (1000000)	(.....)	3

[a]Fitnes je počítaný pomocou Hammingovej vzdialenosti medzi daným fenotypom (foldingom) a požadovaným cieľovým fenotypom, ktorý je vyjadrený p_{req}=(((.))).

Pojem foldingu nám umožňuje zaviesť triádu fundamentálnych koncepcií evolučnej biológie, a to *genotyp*, *fenotyp* a *fitnes*. Genotyp $\boldsymbol{g}$ je reprezentovaný binárnym reťazcom dĺžky n, $\boldsymbol{g} = (g_1 g_2 ... g_n) \in \{0,1\}^n$, fenotyp $p(\boldsymbol{g})$ je priradený foldingu reťazca $\boldsymbol{g}$; formálne, môže byť vyjadrený ako reťazec znakov $p(\boldsymbol{g}) = fold(\boldsymbol{g}) \in \{(,),.\}^n$. Konečne,

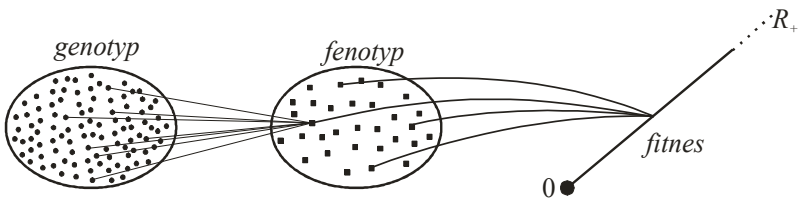
fitnes je numerický atribút $\boldsymbol{g}$, špecifikuje podobnosť reťazca znakov $p(\boldsymbol{g})$ s požadovaným reťazcom znakov p_{req}, ktorý reprezentuje cieľový fenotyp evolúcie systému

$$fitness(\boldsymbol{g}) = s\left(p(\boldsymbol{g}), p_{\text{req}}\right) \tag{23}$$

Maximálny fitnes rovný n majú len tie binárne reťazce $\boldsymbol{g}$, ktorých fenotyp (folding) $p(\boldsymbol{g})$ je totožný s požadovaným foldingom p_{req}, t.j. $fitnes(\boldsymbol{g}) = n$, ak $p(\boldsymbol{g}) = p_{\text{req}}$. V opačnom prípade, ak $s(p(\boldsymbol{g}),p_{\text{req}})<n \Leftrightarrow p(\boldsymbol{g}) \neq p_{\text{req}}$, potom fitnes je menšie ako n; v limitnom prípade, ak folding $\boldsymbol{g}$ je totálne odlišný od požadovaného foldingu, t. j. $s(p(\boldsymbol{g}),p_{\text{req}})=0$, potom fitnes $\boldsymbol{g}$ je nulové, $fitnes(\boldsymbol{g})=0$. Tabuľka 1 obsahuje ilustratívne príklady, kde genotyp je reprezentovaný všetkými možnými binárnymi reťazcami dĺžky 7, ich počet je $128 = 2^7$, tieto reťazce sú ohodnotené fenotypom – foldingom a fitnes. Z tejto tabuľky vyplýva, že mnoho rôznych reťazcov má rovnaký folding, a podobne, mnoho rôznych fenotypov – foldingov je ohodnotených rovnakým fitnes. Môžeme teda formulovať všeobecné pozorovanie, že tieto dve zobrazenia, genotyp na fenotyp a fenotyp na fitnes, sú typu mnoho-na-jeden, pozri obrázok 8. Jednotlivý binárny reťazec môže byť ohodnotený mnohými foldingami (ktoré však majú rovnaký počet spriahnutých elementov), ale zobrazenie produkované Algoritmom 2 je prvé, ktoré obsahuje daný maximálny počet spriahnutí.

Vzájomný vzťah medzi triádou „genotyp-fenotyp-fitnes“ je formálne špecifikovaný dvoma zobrazeniami (pozri obrázok 9)

$$G \xrightarrow{p} P \xrightarrow{fitness} [0,\infty) \tag{24}$$



Obrázok 9. Schematické znázornenie zloženého zobrazenia (24). Obe zobrazenia sú typu mnoho-na-jeden, t.j. mnoho reťazcov je zobrazených na jeden fenotyp – folding, podobne, mnoho fenotypov je zobrazených na jeden fitnes. To znamená, že v týchto zobrazeniach existuje ***veľká redundantnosť*** kódovania genotypu, t.j. mnoho genotypov (binárnych reťazcov) je zobrazených na jeden folding, a taktiež, mnoho fenotypov – foldingov je ohodnotených jedným fitnes. Táto vlastnosť silnej redundantnosti kódovania genotypu je pokladaná za veľmi dôležitú vlastnosť, ktorá umožňuje vznik neutrálnych mutácií a ich zásadného významu pre Darwinovu evolučnú teóriu.

To znamená, že základná entita genotyp je v prvom kroku zobrazená (procesom embryogenézy) na fenotyp, a potom fenotyp (ako dôsledok jeho správania sa v prostredí) je zobrazený (vyhodnotený) na fitnes, to znamená, že formálne existuje ***zložené zobrazenie*** fenotypu priamo na fitnes

$$G \xrightarrow{f = fitness \circ P} [0,\infty) \tag{25}$$

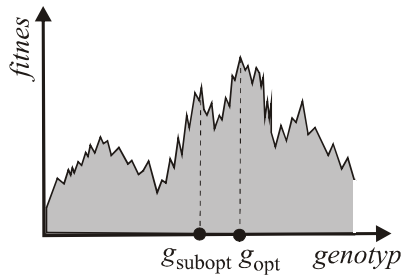
Toto nové zobrazenie mapuje priamo genotyp na fitnes bez nutnosti poznať medzikrok reprezentovaný fenotypom. Musíme však poznamenať, že takýto formálny prístup k vytváraniu zloženého zobrazenia genotypu priamo na fitnes je plne akceptovateľný len z matematického hľadiska. Koncepcia fenotypu v tomto formálnom modeli Darwinovej evolúcie je veľmi plodnou heuristikou. Daná forma fenotypu p_{req} môže byť pokladaná za

evolučný cieľ, na základe čoho môžeme povedať, že Darwinova evolúcia je reprezentovaná sekvenciou fenotypov, ktoré sa progresívne blížia k evolučnému cieľu (reprezentovaného postulovaným fenotypom p_{req}).

Ako už bolo uvedené v úvodnej časti tejto práce, Sewall Wright v r. 1932 [36] zaviedol jeden z fundamentálnych pojmov neodarwinizmu, a to pojem *povrchu fitnes* (pozri obrázok 10). Pomocou tohto pojmu môžeme charakterizovať Darwinovu evolúciu ako optimalizačný proces, kde evolvujúca populácia hľadá globálne maximum (alebo iné riešenie, ktoré je mu blízke)

$$g_{\mathrm{opt}} = \arg\max f(g) \tag{26}$$

Tento komplexný optimalizačný problém bude v ďalšej časti práce riešený použitím metód umelej chémie, ktoré sú založené na metafore Eigenových replikátorov. Bude ukázané, že „chemostat“ poskytuje efektívny prostriedok pre riešenie optimalizačných problémov typu (26), t.j. replikátorové metódy umelej chémie sú vhodné prístupy na simuláciu Darwinovej evolúcie na molekulárnej úrovni.



Obrázok 10. Povrch funkcie fitnes bol pôvodne zavedený ako jeden zo zdrojov moderného neodarwinizmu americkým populačným genetikom Sewallom Wrightom v r. 1932 [36], ktorý charakterizoval Darwinovu evolúciu ako optimalizačný proces prebiehajúci nad fitnes povrchom, ktorý je špecifikovaný zloženým zobrazením (25) f . Výsledný genotyp populácie je priradený bodu $\boldsymbol{g}_{\mathrm{opt}}$, v ktorom má fitnes maximálnu hodnotu. Podľa *motta* tejto práce, Darwinova evolúcia musí obsahovať mechanizmy, ako prejsť zo suboptimálneho riešenia $\boldsymbol{g}_{\mathrm{subopt}}$ na optimálne riešenie $\boldsymbol{g}_{\mathrm{opt}}$.

Pristúpime k sumarizácii výsledkov tejto sekcie. ***Obe zobrazenia genotypu na fenotyp a fenotypu na fitnes sú silne redundantného a stochastického charakteru typu mnoho-na-jeden. Všeobecné vlastnosti týchto zobrazení môžu byť študované pomocou počítačových simulácií, t.j. vlastnosti týchto zobrazení generujeme pomocou počítača a snažíme sa získané výsledky zovšeobecniť pomocou štatistických veličín.***

6 Simulácia molekulárnej Darwinovej evolúcie pomocou chemostatu

Metóda chemostatu, ktorá bola špecifikovaná v prvej časti tejto práce, bude použitá ako algoritmický základ pre simuláciu molekulárnej Darwinovej evolúcie. Chemostat bol inicializovaný identickými reťazcami dĺžky 39

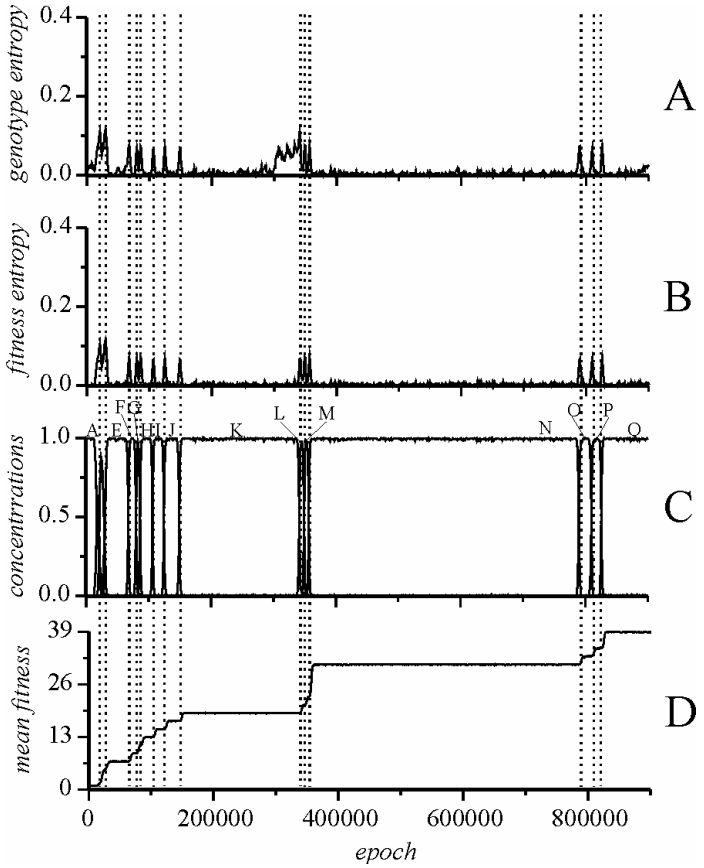
$$\boldsymbol{g}_{\mathrm{ini}}=\mathtt{111111111111111000000000000000000000000} \tag{27a}$$

Požadovaný cieľový fenotyp má tvar

$$p_{\mathrm{req}}=\mathtt{((((((.))))))((((((.))))))((((((.))))))} \tag{27b}$$

Tento folding môže byť názorne interpretovaný pomocou trojlistovej vrtule (pozri sekundárnu štruktúru Q na obrázku 11). Tento fenotyp bude slúžiť ako cieľový, všetky

výpočty fitnes binárnych reťazcov budú uskutočnené vzhľadom k tomuto cieľovému fenotypu – foldingu. To znamená, že očakávame emergenciu takého genotypu v chemostate, ktorého folding je totožný s p_{req} špecifikovaného (27b). Operátor mutácie O_{mut} je určený pravdepodobnosťou 1-bitovej mutácie P_{mut}=0.0001. Použitá pravdepodobnosť je dostatočne malá, aby bola zabezpečená evolučná trajektória obsahujúca blízke reťazce s binárnymi reťazcami, ktorých folding je určený (27b). Ďalší dôležitý parameter chemostatu je jeho veľkosť, ktorá bola nastavená na 500.



Obrázok 11. Priebehy (A) entropie genotypu, (B) entropie fitnes, (C) koncentrácií reťazcov s daným fitnes a (D) priemerné fitnes. Priemerné fitnes má neklesajúci priebeh s relatívne dlhými neutrálnymi stázami. Jej „schodový“ charakter je indikovaný pomocou entropií, náhla zmena priemerného fitnes je sprevádzaná aj dočasným zvýšením entropií. To znamená, že v týchto oblastiach dochádza k premnoženiu „neutrálnych mutantov“, ktoré majú schopnosť sa pomocou 1-bitových mutácií pretransformovať na nový genotyp, ktorého fenotyp je podstatne odlišný od fenotypu pôvodného genotypu. Priebeh (C) znázorňuje relatívne koncentrácie reťazcov s daným aktuálnym fitnes. Každý koncentračný profil je identifikovaný fenotypom z obrázku 12.

Zloženie chemostatu je charakterizované pomocou dvoch typov entropií, ktoré sú určené takto: Nech $w(x)$ je pravdepodobnosť toho, že v chemostate existuje reťazec, ktorý je ohodnotený veličinou $\boldsymbol{x}$ (fitnes alebo genotyp). Potom (pozri ref. [31])

$$S_X = -\sum_x w(x)\ln\left[w(x)\right] \tag{28}$$

Tieto entropické parametre sú veľmi citlivé na zmeny zloženia chemostatu. Pre homogénny chemostat (ktorý obsahuje len rovnaké reťazce) sú obe entropie nulové. Fitnes

entropia S_{fitnes} môže slúžiť ako vhodný indikátor prechodných stavov, kde nové genotypy s vyšším fitnes nahradzujú staršie genotypy s nižším fitnes. Tento moment prechodu je detegovaný pomocou náhlej zmeny fitnes entropie S_{fitnes} z nulovej hodnoty na kladné hodnoty. Genotypová entropia je schopná detegovať také prechodné stavy, kde fitnes zostáva nezmenená, ale zloženie genotypov sa mení.

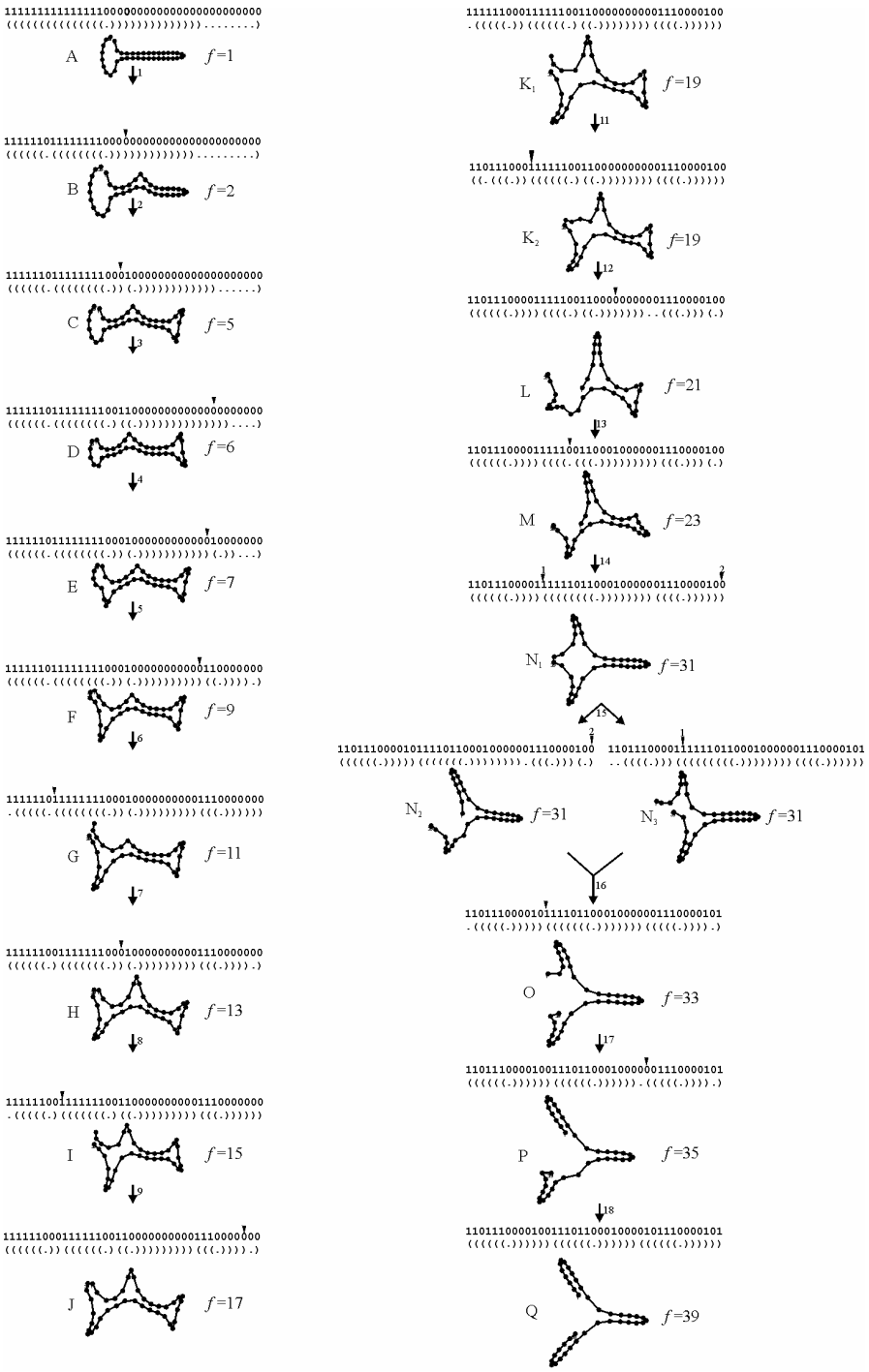
Numerické výsledky našich počítačových simulácií molekulárnej Darwinovej evolúcie sú znázornené na obrázkoch 11 a 12. Vidíme, že priebeh priemerného fitnes v chemostate (diagram D na obrázku 11) je neklesajúca „schodová“ funkcia s niekoľkými relatívne dlhými neutrálnym stázami. Diagramy A a B reprezentujú priebehy genotypovej a fitnes entropie. Tieto priebehy jednoznačne indikujú to, že na konci prvej neutrálnej stázy (približne v 300000. epoche) dochádza k podstatnému odklonu od genotypovej homogénnosti, t. j. v tejto evolučnej etape chemostat obsahuje dva odlišné genotypy s rovnakým fitnes. Tento odklon od homogénnosti chemostatu môže byť chápaný ako nutná podmienka skutočnosti, že nastáva evolučný prechod do nasledujúceho stavu s vyšším fitnes.

Podobná situácia nastáva aj v predchádzajúcich etapách evolúcie, pri fenotypových prechodoch A→B a E→F. Diagram C reprezentuje koncentračné profily rôznych genotypov s daným fitnes. Vidíme, že tieto profily pekne súhlasia s Eigenovou teóriou predikcie koncentračných profilov (pozri obrázok 6, diagram B). Vzniká sekvencia dobre separovaných koncentračných profilov. Bodkované vertikálne pomocné čiary na obrázku 10 dobre indikujú zmeny entropie. Dostali sme 17 rôznych genotypov, ktoré sú ohodnotené rôznymi fenotypmi (foldingmi).

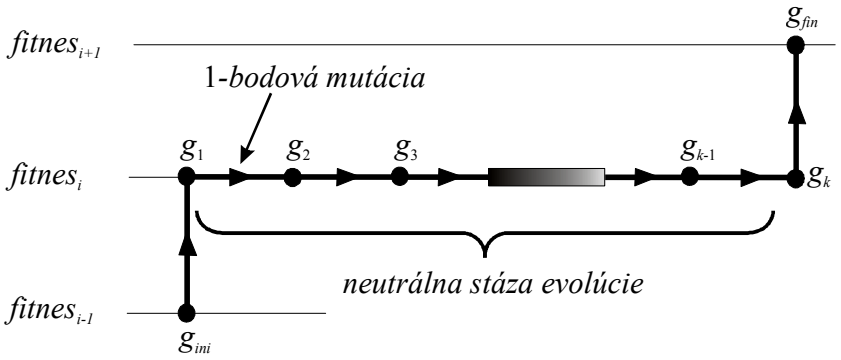
Z týchto výsledkov vyplýva, že neutrálne mutácie hrajú dôležitú úlohu v molekulárnej Darwinovej evolúcii ako mechanizmu (spomínaného Sewallom Wrightom) umožňujúcemu opustiť lokálne maximá na povrchu fitnes. (pozri obrázky 13 a 14). V mnohých prípadoch daný reťazec $\boldsymbol{g}$ nie je „susedným“ reťazcom na evolučnej trajektórii s iným reťazcom $\boldsymbol{g}'$, ktorý by mal vyššie fitnes ako predchádzajúci reťazec (t.j. prechod $\boldsymbol{g} \to \boldsymbol{g}'$ nevytvára evolučný skok). Môžeme povedať, že reťazec $\boldsymbol{g}$ je úspešný v tom, že stochasticky hľadá iný reťazec $\tilde{\boldsymbol{g}}$ v okolí reťazca $\boldsymbol{g}$, ktorý má rovnaké fitnes, ale pre ktorý evolučný skok $\tilde{\boldsymbol{g}} \to \boldsymbol{g}'$ s rastom fitnes, $fitnes(\boldsymbol{g}') > fitnes(\tilde{\boldsymbol{g}})$, už existuje.

7 Súhrn a diskusia

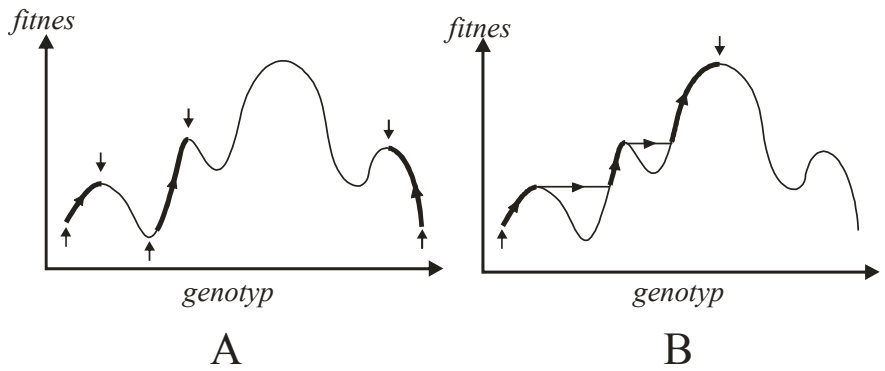
Hlavným cieľom tejto práce je ukázať, že *in-silico* simulácie molekulárnej darwinovskej evolúcie poskytujú detailný pohľad na jej priebeh a mechanizmy. Autoreplikujúce biomakromolekuly v našich simuláciách boli reprezentované pomocou *p*-nárnych reťazcov, pričom ich fitnes bol počítaný pomocou zovšeobecnenej Kauffmanovej funkcie. Tento prístup k *in-silico* simulácii darwinovskej evolúcie bol prednedávnom použitý Newmanom a Engelhardtom [30], ktorí na základe svojich výsledkov dospeli k veľmi dôležitému záveru, že tento jednoduchý „formálny“ prístup simuluje v podstate všetky výsledky získané pomocou fyzikálneho modelu RNA a jej foldingu na sekundárnu štruktúru, ktoré boli dosiahnuté teoretickým chemikom (tiež aj teoretickým biológom) Petrom Schusterom so spolupracovníkmi [14,31,32]. Naše výsledky, pozorovania a zovšeobecnenia môžu byť sumarizované takto:



Obrázok 12. Trajektória $T(\boldsymbol{g}_{\mathrm{ini}}, p_{\mathrm{req}})$, kde počiatočný genotyp je špecifikovaný (27a) (jeho folding je znázornený diagramom A) a cieľový požadovaný fenotyp je špecifikovaný (27b) (diagram Q). Dĺžka tejto trajektórie je 18, vo všetkých prechodových stavoch použité mutácie boli 1-bitové (t.j. medzi danými foldingmi je jednotková Hammingova vzdialenosť). V dvoch prípadoch mutácie majú neutrálny charakter, pri prechodoch 11 a 15. Tieto prechody sú rôzneho charakteru, prvý je špecifikovaný sekvenciou neutrálnych mutácií, zatiaľ čo druhý je špecifikovaný „paralelným“ výskytom dvoch neutrálnych mutácií (t.j. trajektória je rozdelená na dve cesty, ktoré sú spojené v nasledujúcom kroku).



Obrázok 13. Schematické znázornenie dvoch skokov fitnes. V prvej etape sa fitnes skokovo zmení z hodnoty $fitnes_{i-1}$ na novú hodnotu $fitnes_i$. Po tomto skoku nastáva relatívne dlhá neutrálna stáza, kde „trajektória" prechádza pomocou neutrálnych mutácií cez postupnosť genotypov $g_1, g_2,..., g_k$. Potom nastáva skoková zmena fitnes z hodnoty $fitnes_i$ na novú hodnotu $fitnes_{i+1}$. Dĺžka neutrálnej stázy je úmerná Hammingovej vzdialenosti $H(g_1,g_k)$ medzi genotypmi g_1 sa g_k.



Obrázok 14. (A) Ilustratívny priebeh molekulárnej Darwinovej evolúcie keď neexistujú neutrálne mutácie. V tomto prípade evolučná adaptácia končí na najbližšom lokálnom maxime (v tomto prípade môžeme povedať, že Darwinova evolúcia je lokálny optimalizátor, fitnes populácie je „zamrznuté" na najbližšom maxime). (B) V tomto prípade existujú neutrálne mutácie (povrch fitnes obsahuje mnoho „plateau" s konštantnou hodnotou fitnes), lokálny charakter Darwinovej evolúcie môže byť „premostený" neutrálnym „plateau", ktoré umožňujú prekonať údolia s nízkym fitnes. To znamená, že molekulárna darwinovská evolúcia má dobrú šancu opustiť lokálne extrémy a dosiahnuť globálne maximum na povrchu fitnes.

(1) ***Eigenova teória replikátorov*** (2. sekcia) tvorí solídny fyzikálno-chemický základ molekulárnej Darwinovej evolúcie. Riešenie jej diferenciálnych rovníc špecifikovaných plauzibilným výberom rýchlostných konštánt poskytuje časové priebehy relatívnych koncentrácií replikátorov a priemerného fitnes v tvaroch, ktorý je obvykle hypoteticky predpokladaný pre darwinovskú evolúciu. (pozri obr. 3).

(2) Darwinovská evolúcia je komplikovaná interakcia medzi ***monodovskou náhodou a nutnosťou*** [29], medzi deterministickými a stochastickými procesmi, ktoré sú jej integrálnou súčasťou. Obsahuje časti, ktoré sú plne deterministické a predikovateľné (napr. zobrazenie genotypu na fenotyp), ale tiež obsahuje aj časti, ktoré majú silný stochastický charakter a môžu byť špecifikované len použitím štatistickej terminológie

(napr. mutácie). Pravdepodobnosť mutácie nesmie byť veľmi malá, potom by sa zastavila evolúcia, prestali by sa vytvárať „evolučné novinky“, ktorých vhodnosť pre evolúciu je testovaná prirodzeným výberom. V opačnom prípade, ak by pravdepodobnosť mutácie bola veľmi veľká, potom by evolúcia zanikla, zmenila by sa na stochastický slepý algoritmus, ktorý by náhodne „skákal“ po povrchu fitnes. Túto skutočnosť môžeme podľa Monoda formulovať jeho známou vetou „*evolúcia prebieha na hrane chaosu*“.

(3) Idea teoretického genetika *Sewalla* Wrighta o ***povrchu fitnes*** (adaptive landscape) má veľký heuristický význam pre interpretáciu darwinovskej evolúcie a môže byť považovaná za jeden z najväčších teoretických prínosov neodarwinizmu. Podľa tejto idey darwinovská evolúcia môže byť interpretovaná ako špeciálna forma ***evolučného algoritmu*** [17,28] pre riešenie extrémne komplikovaných kombinatoriálnych optimalizačných problémov.

(4) Použitý model poskytuje jednoduchý mechanizmus pre vysvetlenie významu ***neutrálnych mutácií*** pre darwinovskú evolúciu (pozri obrázky 16 a 17). Neutrálne mutácie na povrchu fitnes sú dôležité pre prekonanie „evolučných pascí“ lokálnych maxím. Darwinovská evolúcia obsahuje neutrálne epochy, kde priemerné fitnes zostáva dlhodobo konštantné, ale zloženie reťazcov populácie sa pomaly mení smerom k reťazcom, ktoré sú schopné „skoku“ 1-bitovej mutácie do neutrálneho grafu s vyšším fitnes.

(5) ***Časová orientácia*** darwinovskej evolúcie je jednoznačná a jej základným prejavom je existencia neklesajúceho priemerného fitnes populácie. Reťazce s väčším fitnes sú reprodukované častejšie ako reťazce s malým fitnes, „silnejšie“ reťazce v priebehu evolúcie majú evolučnú výhodu vzhľadom k iným reťazcom. Inými slovami povedané, darwinovská evolúcia je progresívna zmena populačného genotypu, kde priemerný fitnes je neklesajúci v priebehu celej evolúcie.

(6) Dve rôzne časové škály môžu byť rozlíšené v darwinovskej evolúcii, a to adaptívny stav a neutrálny stav. ***Adaptívny stav*** je priradený náhlym zmenám priemerného fitnes (pozri obr. 17), kde koexistujú rôzne fenotypy. Starý fenotyp má menší fitnes ako novovznikajúci fenotyp. Pretože pravdepodobnosť replikácie reťazca je úmerná jeho fitnes, reťazce odpovedajúce novému fitnes sa častejšie replikujú ako staré. Nové reťazce s väčším fitnes zvíťazia v priebehu niekoľkých evolučných krokov, t.j. adaptívny stav, v ktorom sa staré fenotypy vymieňajú za nové fenotypy, sa môže zdať externému pozorovateľovi ako extrémne rýchly evolučný proces. ***Neutrálny stav*** darwinovskej evolúcie spočíva v dlhodobej neutrálnej etape (stáze), kde výskyt neutrálnych mutácií je významný pre stochastickú prípravu nasledujúceho krátkodobého adaptívneho stavu.

(7) Evolučné vlastnosti prezentovaného modelu sú dobre špecifikované pomocou dvoch typov entropie, ktoré boli zavedené k indikovaniu zmien v chemostate. ***Fitnesová entropia*** špecifikuje zmeny v zložení binárnych reťazcov vzhľadom k ich fitnes. Na druhej strane, ***genotypová entropia*** je veľmi citlivá na každú kompozičnú zmenu v chemostate, je schopná postihnúť aj zmeny spôsobené neutrálnymi mutáciami. To znamená, že spoločné sledovanie týchto dvoch entropií v priebehu evolúcie

chemostatu je schopné odlíšiť neutrálne a neneutrálne mutácie na evolučnej trajektórii chemostatu.

(8) Aké sú ***limity prezentovaného modelu*** darwinovskej evolúcie? Použitý tvar genotypu a jeho zobrazenia na fenotyp sú extrémne jednoduché, je vhodný pre štúdium evolúcie u vírusov, bakteriofágov a najjednoduchších prokaryotických baktérií. Pre zložitejšie organizmy musí byť model genotypu (a jeho zobrazenia na genotyp) omnoho zložitejší, musí zahrňovať také koncepty, akými sú premenlivá dĺžka genotypu, hierarchická modulárna štruktúra, iné mutácie než aké sú 1-bodové mutácie, atď. Jeden z najdôležitejších problémov súčasnej teórie Darwinovej evolúcie je vysvetlenie a emergencia tzv. ***neredukovateľných zložitostí*** (***irreducible complexities***) [3]. Použitý teoretický model Darwinovej evolúcie nie je schopný vysvetliť tento problém ani na elementárnej úrovni. Súčasné aktivity v umelom živote sa okrem iného usmerňujú aj na oblasti, ktoré by mohli mať význam pre vypracovanie všeobecnej teórie Darwinovej evolúcie. Veľmi intenzívne sú študované modulárne aspekty genotypu [22,23] a problém symbiózy [21], kde oba problémy vyžadujú zložitejšie genotypy ako ten, ktorý bol študovaný v tejto práci. Tieto a podobné problémy sú podrobne študované v knihe mladého anglického bioinformatika R. A. Watsona [35].

Poďakovanie: Autor ďakuje Jirkovi Pospíchalovi za trpezlivosť pri sledovaní a rozvíjaní mojich „endless“ kontemplácií o Darwinovej evolúcii. Táto práca bola podporená grantami 1/4053/07 Vedeckej grantovej agentúry SR a APVV 20-002504.

Literatúra

[1] Atkins, P. W.: *Physical Chemistry*. Oxford: Oxford University Press 1998.
[2] Banzhaf, W., Dittrich, P., Eller, B.: Topological Interactions in a Binary String System. *Physica D* **125** (1999) 85-104.
[3] Behe, M. J.: *Darwin's black box*. New York: Simon & Schuster 1996.
[4] Banatre, J.-P., Le Metayer, D.: The Gamma Model and its Discipline of Programming. *Sci. Comp. Progr.* **15** (1990) 55-77.
[5] Berry, G., Boudol, G. : The Chemical Abstract Machine. *Theoret. Comp. Sci.* **96** (1992) 217-248.
[6] Bobrík, M., Kvasnička, V., Pospíchal, J.: Artificial Chemistry and Molecular Darwinian Evolution of DNA/RNA-Like Systems. In Kelemen, A., Abraham, A., Chen, Y. *Computational Intelligence in Medical Informatics*. Berlin: Springer 2007 (in print).
[7] Dawkins, R.: Universal Darwinism. In Bendall, D. S. (ed.) *Evolution from Molecules to Men*. London: Cambridge Univesity Press 1982, p. 403.
[8] Dennett, D. C.: *Darwin's Dangerous Idea - Evolution and the Meaning of Life.* London: Penguin Press 1995.
[9] Dittrich, P.: *Artificial Chemistries* (tutorial material). A tutorial held at *ECAL'99*, 13-17 September 1999, Lausanne, Switzerland.
[10] Eigen, M.: Self organization of matter and the evolution of biological macro molecules. *Naturwissenschaften* **58** (1971) 465-523.

[11] Eigen, M., Schuster, P.: The Hypercycles: A Principle of Natural Evolution. *Naturwissenschaften* **64** (1977) 541-565; **65** (1977) 7-41; **65** (1977) 341-369.
[12] Fogel, D. B.: *Evolutionary Computation. Toward a New Philosophy of Machine Intelligence*. New York: The IEEE Press 1995.
[13] Fonatana, W. and Schuster, P.: Continuity in Evolution: On the Nature of Transitions. *Science* **280** (1998) 1451-1455.
[14] Fontana, W.: Algorithmic Chemistry. In: Langton, C. G. (ed.) *Artificial Life II.* Reading, MA: Addison Wesley 1991, p. 159.
[15] Gálik, D.: Univezálny darvinizmus (Môže byť darvinizmus teóriou všetkého?), Zborník *Kognitívne vedy III (CogSci 2000)*, Katedra matematiky CHTF STU, Bratislava (prednáška je dostupná na http://math.chtf.stuba.sk/ CogSci_2000.html)
[16] Gillespie, D. T.: Exact Stochastic Simulation of Coupled Chemical Reactions. *J. Phys. Chem.* **81** (1977) 2340-2361.
[17] Holland, J. H.: *Adaptation in Natural and Artificial Systems*. Ann Arbor: University of Michigan Press 1975.
[18] Jones, B. L., Enns, R. H., Rangnekar, S. S.: On the theory of selection of coupled macromolecular systems. *Bulletin of Mathematical Biology* **38** (1976) 15-28.
[19] Kauffman, S. A.: *The Origins of Order: Self-Organization and Selection in Evolution*. New York: Oxford University Press 1993.
[20] Kauffman, S. A., Hohnsen, S.: Coevolution to the edge of chaos: Coupled fitnes landscape, poised states, and coevolutionary avalanches. *J. Theor. Biology* **149** (1991) 467-506.
[21] Kvasnička, V.: An evolutionary model of symbiosis. In: P. Sinčák, J. Vasčák (eds.): *Quo Vadis Computational Intelligence?* Heidelberg: Physica-Verlag 2000, p. 293.
[22] Kvasnička, V.: An evolutionary simulation of modularity emergence of genotype-phenotype mappings. *Neural Network World* **5** (2001) 473-491.
[23] Kvasnička, V.: A modularity emergence of genotype-phenotype mappings. *Artificial Life* **8** (2002) 295-310.
[24] Kvasnička, V., Pospíchal, J.: A study of autoreplicators and hypercycles by typogenetics. In: J. Kelemen, P. Sosik: *Advances in Artificial Life*, ECAL 2001, LNAI 2159, Springer 2001, p. 37.
[25] Kvasnička, V., Pospíchal, J.: Autoreplicators and Hypercycles in Typogenetics, *Journal of Molecular Structure (Theochem)* **547** (2001) 119-138.
[26] Kvasnička, V., Pospíchal, J.: Artificial Chemistry, Replicators, and Molecular Darwinian Evolution *In Silico.* In: P. Sinčák, J. Vasčák (eds.): *Quo Vadis Computational Intelligence?* Heidelberg: Physica-Verlag 2002, p. 403.
[27] Kvasnička, V., Pospíchal, J.: Artificial Chemistry and Molecular Darwinian Evolution in silico. *Collection of Czechoslovak Chemical Communication* **68** (2003) 139-177.
[28] Kvasnička, V., Pospíchal, J., Tiňo, J.: *Evolučné algoritmy.* Bratislava: Vydavateľstvo STU 2000.
[29] Monod, J.: *Chance and Necessity. An Essay on the Natural Philosophy of Modern Biology.* New York: Vintage 2000.
[30] Newman, M. E. J., Engelhardt, R.: Effects of neutral selection on the evolution of molecular species. *Proc. Roy. Soc. London B* **265** (1998) 1333-1338.

[31] Schuster, P.: Molecular Insight into Evolution of Phenotypes. In *Evolutionary Dynamics - Exploring the Interplay of Accident, Selection, Neutrality, and Function*, ed. by Crutchfield, J. P. and Schuster, P., Oxford University Press 2002.

[32] Schuster, P., Fontana, W.: Chance and Necessity in Evolution: Lessons from RNA. *Physica D* **133** (1999) 427-452.

[33] Spiegelman, S.: An Approach to the Experimental Analysis of Precellular Evolution. *Quart. Rev. Biophysics* **4** (1971) 213-253.

[34] Tomáška, Ľ., Nosek, J.: Molekulárna darwinovská evolúcia: In vivo, in vitro a in silico. Zborník *Kognitívne vedy III (CogSci 2000)*, Katedra matematiky CHTF STU, Bratislava (prednáška je dostupná na http://math.chtf.stuba.sk/cogsci_2000.htm).

[35] Watson, A. R.: *Compositional Evolution. The Impact of Sex, Symbiosis, and Modularity on the Gradualist Framework of Evolution*. Cambridge, MA: MIT Press 2006.

[36] Wright, S.: The Roles of Mutation, Inbreeding, Crossbreeding, and Selection in Evolution. *Proceedings of The Six International Congress of Genetics,* (1932). Vol. I, 356. (http://www.blackwellpublishing.com/ridley/classictexts/wright.pdf).

Ontogeneze jako algoritmus a jako výklad

Anton Markoš, Lenka Sýkorová[1], Libor Sýkora[2]

Abstrakt. Ontogenetické procesy jsou často líčeny jako realizace hotového, evolucí vyladěného programu zapsaného v DNA. Nabízíme alternativní pohled, kdy genetický zápis je čtením, nikoli pouhým dekódováním.

Každý akt čtení je tedy složitou transakcí mezi kompetencí čtenáře (jeho poznatky o světě) a charakterem kompetence, kterou postuluje daný text, tak aby se dal číst úsporným způsobem.

U. Eco [3]

Ruský sémiotik Jurij Lotman [2] konstatuje, že každý text (textem budeme rozumět sled znaků z konečné množiny prvků) je zprávou, která se nabízí na dva způsoby: jako něco, co lze dekódovat pomocí předem daných tabulek kódů, nebo jako zpráva určená ke čtení. První případ vyžaduje použití umělých jazyků, které z jazyka přirozeného přebírají pouze jednu z jeho funkcí: schopnost předat zprávu. Výhodou je jistá jednoznačnost přenosu zprávy a možnost celý proces mechanizovat. V zájmu zvládnutí této funkce se však ze sémiotických struktur odstraní jejich potenciál sloužit dalším funkcím, které za přirozených podmínek jsou přítomny. Především jde o funkci tvůrčí – systém přirozeného jazyka není jenom přenašečem hotových zpráv, ale také generuje zprávy nové.

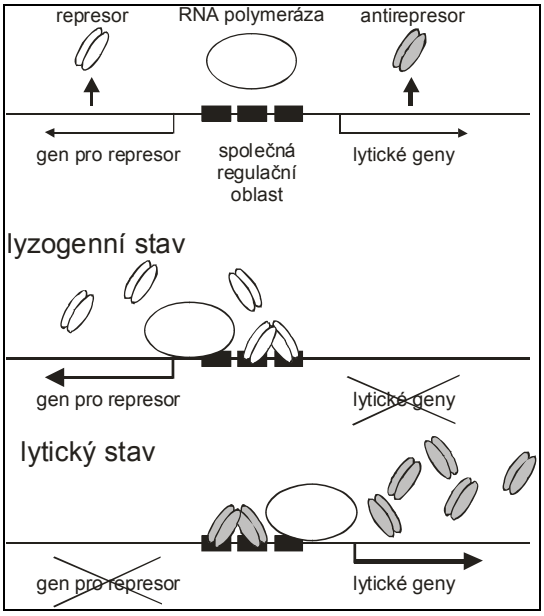
Opačným pólem umělých jazyků jsou sémiotické systémy se zdůrazněnou tvůrčí složkou. Zde překladatel musí *vybírat* jednu z možných interpretací textu. Překlad zapsaného se stává tvůrčím aktem: nikoli dekódováním, ale výkladem. Jen v tomto druhém případě se mohou objevovat nové, nečekané výklady textu a nová řešení. Přirozený jazyk potřebuje mluvčího, protože jej nelze uzavřít do tabulek kódů; tím však – protože neexistují dva stejní mluvčí – se v procesu čtení generuje bezpočet výkladů téhož textu. Soubor mluvčích pak v rámci dané kultury (sémiosféry) stanoví závazné, kanonické čtení – soubor takových interpretací daných textů, které budou pro komunitu akceptovatelné. Z toho plyne to, že ne všechny možné interpretace se mohou uplatnit, a také to, že případné nové výklady už budou ovlivněny předchozími. Plyne však z toho také to, že tento kanonický výklad není poznatelný na úrovni samotného zápisu, je výsledkem hermeneutického úsilí komunity a jako takový se uchovává jinak než v zápisu. Jazyk předchází sdělením, která jsou v něm vyslovena. Dokladem toho, říká Lotman, mohou být novátorské počiny v umění, které často způsobí šok a pohoršení jen proto, že na rozdíl od autora auditorium neumí sdělení interpretovat, a to navzdory společné jazykové tradici autora i publika. Po jisté době, kdy se publikum s danou formou sžije, může převládnout i v tomto případě formalizovaný metajazyk, který generuje sdělení automaticky a tato sdělení jsou také jako kódy přijímána. Existuje tedy spektrum textů na

---

[1] Přírodovědecká fakulta UK Praha, E-mail: markos@natur.cuni.cz
[2] Biometrics R&D s.r.o.
[3] Eco, U. *Medzi autorom a textom*, ze sborníku [1] str. 70.

ose umělý – umělecký, avšak krajní póly v reálných jazycích neexistují v čisté formě. Jakmile vznikne formalizovaná forma zápisu, je nutné počítat s apriorními znalostmi interpreta a tím i s historií.

Pokusme se Lotmanův příměr přenést do oblasti ontogeneze mnohobuněčných organismů. Vždyť ontogeneze se v dnešní době modeluje právě jako interakce textu uloženého formou posloupnosti nukleotidů v DNA s interpretačním zařízením, které představuje buňka či mnohobuněčné tělo. V jakém modu tato interpretace probíhá – v modu dekódovacím nebo výkladovém?



Obrázek 1. *Schéma regulace projevů bakteriofága λ.* Horizontální čára představuje molekulu DNA s vyznačenými oblastmi genů (lomené šipky) a *cis*-regulační oblasti (tmavé bloky). RNA polymeráza je enzym, který přepisuje protein kódující části zápisu do RNA. Enzym přepisuje buď geny vpravo nebo vlevo, podle toho, kam je mu dovoleno se na DNA navázat, tj. podle toho, které ze dvou možných míst je volné, neobsazené *trans*-regulátory. Produktem přepisovaných genů je řada proteinů, z nichž dva – označené jako represor a antirepresor, jsou *trans*-regulátory, které se váží na regulační oblast. Fág má dva mody existence, podle toho, zda se čte levá nebo pravá část genetického zápisu. V lytickém stavu se fág množí a zabíjí hostitele, v lyzogenním stavu je latentní součástí genomu a kromě molekuly represoru se syntetizují pouze proteiny zajišťující imunitu bakterie vůči infekci jinými fágy. Oba stavy se navzájem vylučují; v lyzogenním stavu represor (*trans*-faktor) vazbou na *cis*-regulační oblast brání spuštění lytické větve genomu a naopak aktivuje lyzogenii. Teprve odstranění represoru (vnějšími faktory) dovolí spuštění lytických genů, mezi nimi i genu pro antirepresor, který je *trans*-faktorem blokujícím lyzogenní aktivity. Silně zjednodušeno.

Oba pohledy mají dlouholetou tradici, zde se ale omezíme na několik posledních desetiletí, v průběhu kterých výrazně převažoval názor, že morfogeneze by se dala vysvětlit jako dekódovací proces. V počátcích tohoto vývoje stály dva milníky:

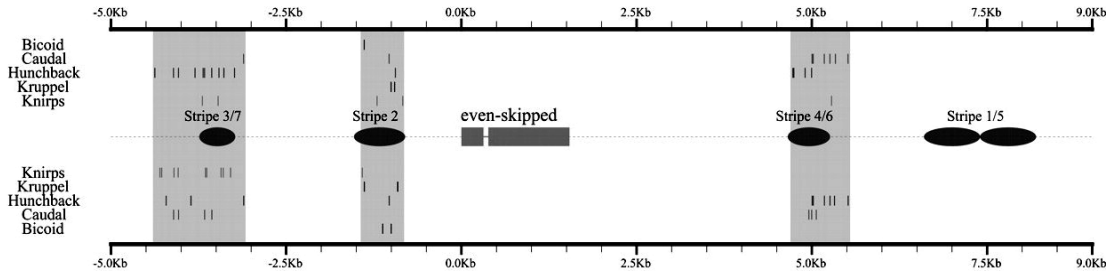
- Odhalení genetického kódu, tj. pravidla, s jakým se v procesu translace přikládají podle posloupnosti nukleotidů v mRNA aminokyseliny do rostoucího

polypeptidového řetězce proteinů. Celý proces používá neměnné postupy, tripletu AUG bude odpovídat aminokyselina metionin, UUU fenylalanin, atd.

- Vypracování operonového modelu regulace genové exprese, tj. regulačních obvodů umožňujících řízené zapínání a vypínání syntézy mRNA (obr. 1). Množina mRNA tedy určuje, jaké proteiny bude buňka vyrábět, a množina proteinů (transkripčních faktorů, *trans*-regulátorů) rozhoduje svou vazbou na *cis*- regulační místa na DNA, které geny budou (v následujícím kroku) přepisovány do RNA. Nás budou v souvislosti s morfogenezí zajímat právě tyto algoritmy. To proto, že zatímco *cis*-oblasti na DNA se mohou měnit pouze mutací, *trans*-regulátory mohou své vlastnosti měnit v širokém rozmezí, v závislosti na množině přítomných regulátorů, na fyzikálních i sémantických faktorech prostředí, a také na historii systému.

V průběhu let se ukázalo, jak složitou síť tvoří genové regulace, zejména u mnohobuněčných organismů, o které nám zde v souvislosti s morfogenezí jde. Už v případě vztahu gen-protein(y) je vztah mezi pořadím nukleotidů v genu a výsledným proteinem víc než volný a musí být spoludefinován celou řadou faktorů epigenetických, tj. nesouvisejících s informací zapsanou v genu.

Výzkum kódu druhého řádu (tj. *cis-trans* interakcí) záhy odhalil řadu dvojic *cis*-motivů a jim odpovídajících transkripčních regulátorů (viz např. [3])



Obrázek 2. *Oblast cis-regulačních elementů v okolí genu even-skipped.* Gen je exprimován v embryu jako 7 oddělených pásů (stripe); v každém z nich je exprese vyvolána jinou kombinací transkripčních faktorů. Středová čára je schematicky DNA se škálou v kilobázích (Kb). Gen je znázorněn jako 2 bloky (exony) začínající na souřadnici 0.0. Regulační oblasti před i za genem obsahují nakupení *cis*-elementů pěti regulačních genů (jména vlevo) – znázorněna lokalizace pro obě vlákna DNA – do těchto míst se proto přednostně váží příslušné proteiny a dohromady konstituují regulační jednotky. K expresi genu je nutná přítomnost jedné takové jednotky. Podle [3].

Tato informace pak umožnila pomocí řady metod, často odvozených od kryptografické analýzy, odhalit regulační motivy pro jednotlivé geny. Ukázalo se, že množina těchto regulačních motivů je často velká, rozprostírá se napříč rozsáhlými úseky genomu, a nikdy nelze s určitostí tvrdit, že ji známe celou (viz například [4]). Ukázalo se také, že jemné ovlivňování konformace regulačních proteinů (alosterické i kovalentní) a hlavně vzájemné ovlivňování proteinů vyskytujících se ve velkých souborech často jdoucích do desítek, je zdrojem tak vysoké variability možných nastavení, že prakticky nelze pořídit konečný výčet možných stavů sítě a jim odpovídajícího chování. Sofistikované modely (např. NK modely založené na booleovských sítích, [5]) sice

umožňují analýzu chování podobných regulačních sítí, ale jejich aplikace na konkrétní situace nepostoupila příliš od triviálních případů, které lze zvládat i bez modelování. Kromě toho řadu morfogenetických procesů lze modelovat i s pomocí Turingových reakčně-difúzních modelů [5].

Oba typy kódů (kód genetický i cis-trans) jakoby nebyly úplné, jakoby představovaly jen hrubou stavbu pro dosud neodhalené krajkoví a štukaturu kódů třetího, čtvrtého, n-tého řádu. Většina badatelů se dnes snaží odhalit tyto předpokládané kódy.

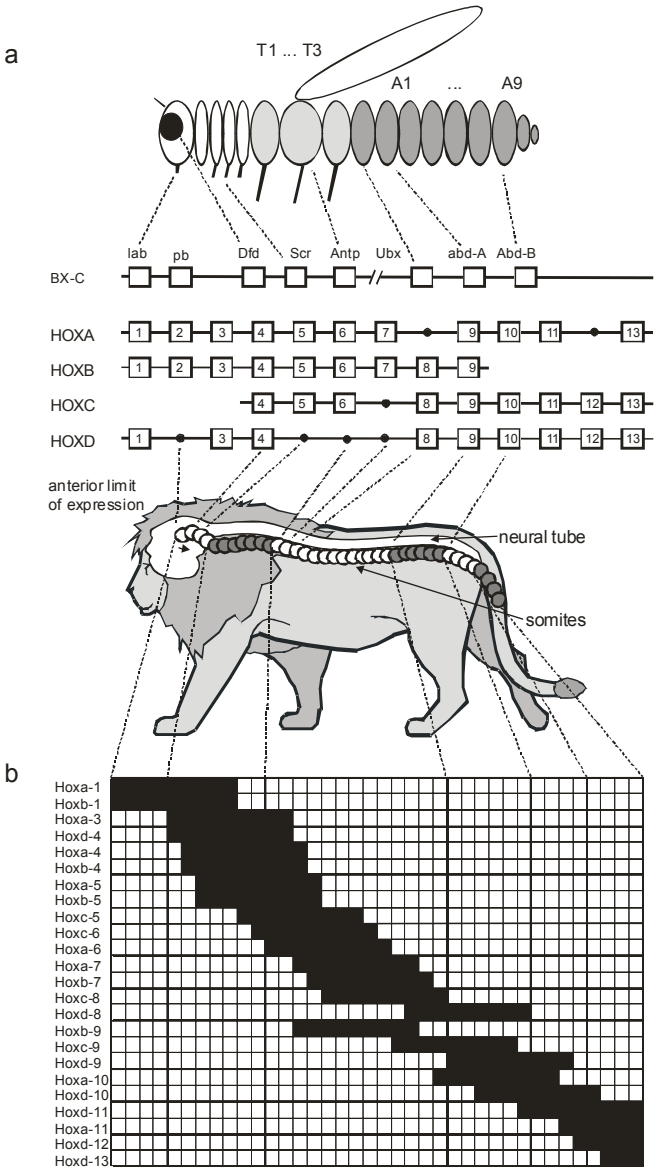
V souvislosti s podobnými programy se vynořuje otázka příčinnosti probíhajících regulačních i stavebních dějů. Vzhledem k tomu, že žádný vývojový proces nemá jasně definované počáteční podmínky (nelze definovat „čas nula“, který by odpovídal „zapnutí počítače“, počátečním a okrajovým podmínkám), je tato otázka velmi závažná, navzdory dnes převládajícímu přesvědčení, že příčinnost je jasná (předpoklad „geny řídí“: geny a jejich epistatické interakce jsou prvotné, funkce a morfologie druhotná). Vliv minulosti a konkrétního prostředí je nesporný – i zde však existuje přesvědčení, že se jednou podaří tyto vlivy formalizovat do vztahů logických operátorů. Povšimněme si, že celý trend výzkumu jako kdyby toužil po tom zbavit se tělesnosti a celý proces modelovat pouze jako virtuální síť vztahů. Samotná interpretace této sítě – tj. výstavba těla – jakoby byla něčím podružným a nezajímavým.

Než představíme alternativu, vložíme dva příklady, které nám pomohou osvětlit rozdíl mezi výše představeným algoritmickým modelem a modelem výkladovým.

1. Role Hox-genů při organizování předozadní osy těla živočichů je notoricky známým a neustále uváděným učebnicovým příkladem (obr. 4). Vyvíjející se embryo dospěje v jedné chvíli do stádia, kdy je tělo předozadně rozčleněno na dobře definované oblasti (v případě obratlovců a členovců jsou to segmenty, ale podobný princip se uplatňuje i u živočichů nečlánkovaných). V každé takto vymezené oblasti dojde k expresi jednoho nebo více Hox genů; tyto geny jsou na chromozomu umístěny za sebou a jejich exprese v těle je většinou kolineární s pořadím na chromosomu. Existuje přísná a většinou jednoznačná korelace mezi strukturou, která se na daném místě vyvíjí, a přítomností produktu(ů) homeotických genů. Příkladem může být moucha octomilka (obr. 4a nahoře). Zjednodušeně: produkt genu *Antp* se vyskytuje ve druhém hrudním článku (T2) a jeho přítomnost koreluje s výskytem páru křídel a 2. páru nohou. Srovnejme teď, co se stane, když se tento gen aktivuje na místech, kde normálně k aktivaci nedochází.

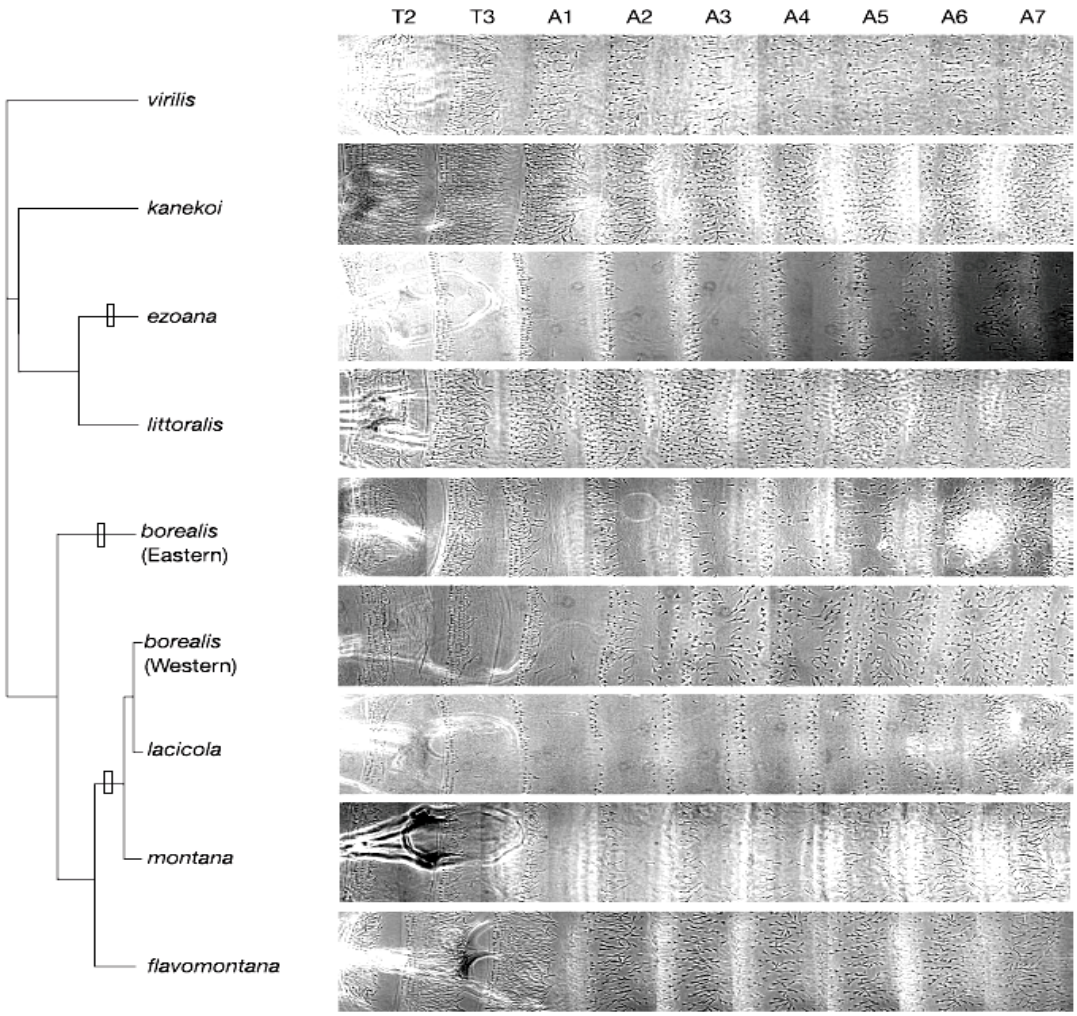
V prvním případě jde o jakýsi záskok: Produkt genu *Ubx* se nachází mj. v článku T3 a jeho výskyt koreluje s tvorbou tzv. halter (deriváty 2. páru křídel) a 3. páru končetin. Nefunkční mutace genu *Ubx* koreluje s rozvratem normální morfogeneze v tomto článku. Obvykle se situace vyřeší tak, že článek imituje situaci v článku předcházejícím – aktivuje se tedy *Antp* a na 3. článku se objeví křídla a ještě jedna sada 2. páru končetin. Jde tedy o příklad vývoje nemístných (ektopických) struktur a s nimi spojené ektopické exprese genů.

Druhý příklad se týká ektopické exprese *Antp*a v přední části hlavy, kde se normálně neaktivují žádné Hox geny. Nemístná exprese genu *Antp* v segmentu nesoucím tykadla koreluje s končetinami vyrostlými místo tykadel. Homeotických mutací obou typů byla popsaná celá řada, a to nejen u hmyzu.



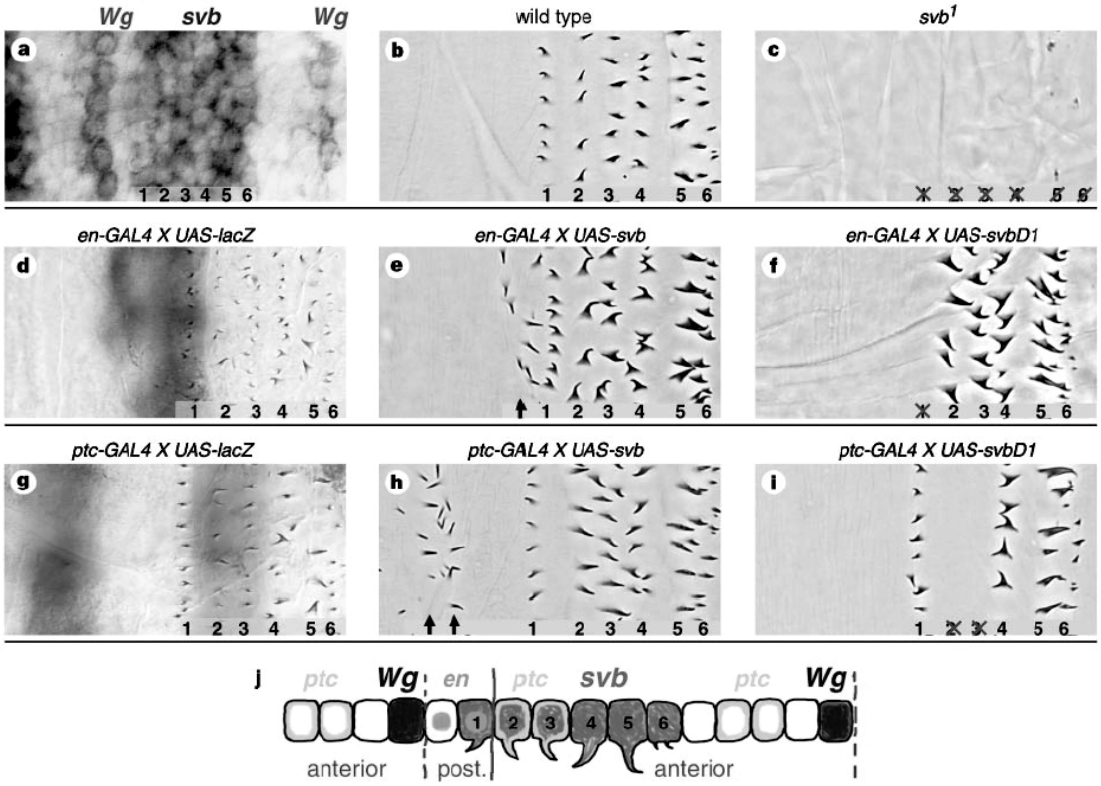
Obrázek 4. *Hox geny a definice předozadní osy živočicha* **a.** Exprese Hox genů podél předozadní osy. Hmyz má jednu sadu těchto genů kolineárních s tělní osou, savci mají sady čtyři. **b.** „Homeotický kód“ – definice tělních článků savce, počínaje prodlouženou míchou. Převzato z [6].

2. Od Hox genů nyní přesuňme pozornost na zdánlivě marginální oblast morfologií u skupiny blízce příbuzných druhů octomilek – často však jde o významný diagnostický znak, umožňující tyto druhy rozpoznat. Z epidermu vyrůstají na jednotlivých segmentech embrya, larvy i dospělce výrůstky – trichomy (denticles). Jejich rozložení je typické pro jednotlivé články, a je také druhově specifické. Na obr. 5 je fylogram skupiny a současně obrázek rozložení trichomů na těle larvy příslušného druhu. Z obrázku je patrné, že ve skupině došlo nezávisle ve třech liniích k silné redukci tvorby trichomů; larvy mají místo trichomů souvislé hladké plochy povrchů (obr. 5).



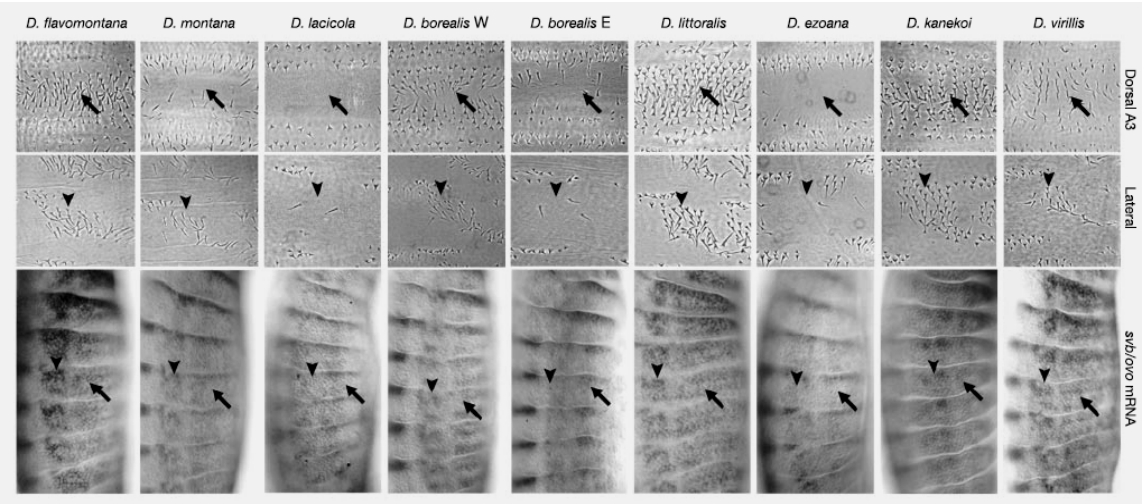
Obrázek 5. Rozložení trichomů na hřbetní straně larev u blízce příbuzných druhů rodu *Drosophila*. Vlevo fylogenetický strom příbuznosti těchto druhů, s vyznačenými událostmi vedoucími ke ztrátě trichomů. Hlava je vlevo, T jsou hrudní, A břišní segmenty. Převzato z [7].

Z předchozího studia tvorby trichomů u *D. melanogaster* se vědělo, že schopnost vytvářet trichomy získají buňky epidermu tím, že se v nich aktivuje gen *ovo/svb* (jeho produkt je transkripčním regulátorem). Mutanti s nefunkčním genem mají hladkou epidermis (obr. 6c). Příklady v obr. 6 dokazují úzkou provázanost exprese genu *ovo/svb* s výskytem trichomů.



Obrázek 6. Korelace aktivity genu *ovo/svb* s obrazcem trichomů. Každé pole obrázku znázorňuje preparát jednoho a téhož abdominálního článku embrya. Schéma dole ukazuje, že parasegment (čárkované hranice) sestává z 12 řad buněk. Exprese genů *wg*, *en* a *ptc* slouží jako markery pro orientaci. Za normálních okolností se gen *svb* aktivuje v řadách označených 1-6; s aktivací koreluje tvorba trichomů, v každé řadě s poněkud jinou morfologií. **a.** Detekce proteinu Wg a Svb; **b.** Rozložení trichomů u normálního embrya; **c.** u mutanta v genu *svb* (od toho název mutace a posléze i genu: *svb* je zkratka od *shaven baby*). Následují příklady transgenních organismů: **d** a **g** jsou konstrukty genů *en* a *ptc* s genem pro enzymy, jejichž přítomnost se projeví ve tkáni zbarvením: tam, kde se exprimuje *en*, resp. *ptc*, se objeví také zbarvení (srovnej se schématem vespod). **e, h**: k těmto konstrukcím je připojena *cis*-regulační oblast (UAS) genu *svb* následována tímto genem. **e:** V důsledku toho se *svb* exprimuje všude tam, kde i *en* – tedy ektopicky i v řadě, předcházející řadě 1. **h:** Totéž pro gen *ptc*: objeví se 2 nové řady trichomů. **f, i:** Podobně jako v předchozím případě, ale konstrukt obsahuje mutovaný nefunkční gen *svbD1*, jehož produkt blokuje účinek *svb* exprimovaného z normálně přítomného genu. Situace je reverzní oproti **e** a **h**: trichomy chybí v řadě 1 (koexprese inhibitoru s *en*), resp. 2 a 3 (koexprese s *ptc*). Podle [8].

Studie [7] srovnávala proto expresi genu ovo/svb u výše zmíněných příbuzných druhů octomilky. Výsledky opět ukazují (obr. 7), že struktury jsou vždy v korelaci s expresí zmíněného genu:



Obrázek 7. Korelace mezi expresí *svb* a rozložením trichomů. Spodní řada: preparáty embryí ze strany, hřbetní část vpravo. Intenzita zbarvení odpovídá expresi *svb* (šipky). Horní dvě řady pak ukazují trichomy na hřbetní a boční straně segmentu, který je vyznačen příslušnými šipkami dole. Šipky označují plochou ekvivalentních oblastí: je vidět, že hustota trichomů koreluje s expresí genu. Převzato z [7].

Jak můžeme výsledky interpretovat? Korelace mezi expresí genu (přítomností proteinu) ovo/svb a morfologií je zajisté dobrá a dala by se znázornit takto:

Protein Ubx — černá skříňka — segment T3

nebo

Protein Ovo/svb — černá skříňka — trichomy

Dokážeme dopodrobna popsat procesy a morfologie na obou stranách černé skříňky, neumíme ale odpovědět na otázku, jak jsou tyto strany provázány příčinně. Většina badatelů bude věřit v příčinnost ve směru zleva doprava, tedy například

Protein Ovo/svb → černá skříňka → trichomy

Absence přímé provázanosti se bude pokládat za nedostatek pouze dočasný, než se podaří najít trik, jak se do skříňky vlámat a odhalit kód či algoritmus, jak je povel transformován v tvar. Podle této představy je to výskyt proteinu (Ubx nebo ovo/svb, vyvolaný předchozí sekvencí programových kroků), který vnutí segmentu příslušnou morfologii. Pokud jsou v pořádku i následné kroky algoritmu, které, jak věříme, existují v černé skřínce, pak v segmentu, kde se vyskytuje protein Ubx, *nemůže nevzniknout* morfologie T3, jakož i kutikula segmentu se *nemůže neobrvit*, navíc druhově a segmentově specifickým způsobem.

Přinejmenším do doby, kdy se tak stane, nabízíme alternativní vysvětlení se šipkami otočenými zprava doleva nebo oboustrannými. Nabízíme hypotézu, že je to naopak tělo, které *za účelem* realizace druhově specifické morfologie volí, spíná expresi genu, tj. volbu potřebných nástrojů. V této optice pohledu je to tělo, které buduje druhově specifickou morfologii, segment T3 *ví*, že mu přísluší budovat morfologii s halterami a 3. párem nohou, a také *ví*, že k tomu potřebuje modul nástrojů jehož součástí je i protein Ubx. Teprve když se – v důsledku mutace – nepodaří tento modul aktivovat, vezme segment zavděk nouzovým řešením a okopíruje chování předchozího segmentu. Podobně

segment – třeba A2 – *ví*, že má/nemá nést soubor trichomů rozložených tak a tak: v místech tohoto rozložení pak opět aktivuje modul obsahující mj. *ovo/svb*, který je z nějakých důvodů k této činnosti potřebný. Není-li protein Ovo/svb k dispozici, zůstane kutikula neozdobena, ať už je to proto, že jde o jev pro druh typický jako u *D. lacicola*, anebo jde o funkční deleci u druhu jinak trichomy tvořícího. Problém s tímto *výkladovým* způsobem morfogeneze je samozřejmě v tom, že nelze ukázat na součásti, kroky algoritmu, kódy, protože ty nikde nejsou zapsány. Toho jsme si vědomi: probíhá-li morfogeneze v módu čtecím, pak k jejímu studiu jsou více příhodné vědy „měkké“ jako je sémiotika nebo hermeneutika.

Literatura

[1] Collini, S., ed.: *Interpretácia a nadinterpretácia*. Bratislava: Archa 1995.

[2] Lotman, Y. M.: *Vnutri mysljaščich mirov*. Moskva: Jazyki russkoj kuľtury 1996.

[3] Berman, B. B., et al.: Exploiting transcription factor binding site clustering to identify cis/regulatory modules involved in pattern formation in the *Drosophila* genome. *PNAS* **99** (2002) 757-762.

[4] Nobrega, M., et al. Scanning human gene deserts for long/range enhancers. *Science* **302** (2003) 413.

[5] Kauffman, S. A.: The *origins of order. Self-organization and selection in evolution.* New York: Oxford University Press 1993.

[6] Markoš, A.: *Readers of the book of life. Conceptualizing evolutionary developmental biology.* Oxford: Oxford University Press 2002.

[7] Sucena, E., et al. Regulatory evolution of *shavenbaby/ovo* underlies multiple cases of morphological parallelism. *Nature* **424** (2003) 935-938.

[8] Payre, F., et al. *ovo/svb* integrates Wingless and DER pathways to control epidermis differentiation. *Nature* **400** (1999) 271-275.

Evolúcia jazyka a univerzálny darwinizmus

Jiří Pospíchal, Vladimír Kvasnička[1]

Abstrakt. V rámci multiagentového systému využívame paradigmu darwinovskej evolúcie. Fitnes v rámci darwinovskej paradigmy v našich modeloch odpovedá schopnosti dorozumieť sa. Tento prístup je analyzovaný z pohľadu Eigenových replikátorov, kde fitnes replikátora odpovedá veľkosti jeho rýchlostnej konštanty replikácie. Modelujeme emergenciu komunikácie dohadovaním sa agentov o priradení signálov významu, kde pokročilejšie modely pomocou započítania vplyvu šumu vysvetľujú výhodu vzniku slov a elementárnej gramatiky.

1 Úvod

V priebehu evolúcie vznikol u živočíchov taký nervový systém – mozog, ktorý im umožňuje pozorovať okolitý svet, učiť sa rôzne typy správania a komunikovať medzi sebou. Jeden druh tejto živočíšnej ríše vytvoril efektívny komunikačný systém, ktorého vyjadrovacie schopnosti sú neohraničené. Ľudský jazyk má túto vlastnosť neohraničenej vyjadrovacej schopnosti preto, že pomocou konečného počtu slov (ktoré vznikajú kombináciou konečného počtu foném) je schopný zostrojiť nekonečný počet viet. Jazyk umožňuje neohraničený prenos negenetickej – kultúrnej informácie medzi jednotlivcami, a tak indukuje nové módy darwinovskej evolúcie. Jazyk umožňuje vznik kultúrnej evolúcie, ktorá ďaleko prekračuje možnosti evolúcie nehovoriacich živočíchov.

Medzi hlavné evolučné inovácie [38] patrí vznik nukleových kyselín, buniek, chromozómov, mnohobunkových organizmov, nervový systém, mozog a jazyk, ktorý však vznikol len u jedného druhu (zatiaľ). Ľudia a humanoidné opice sa evolučne oddelili pred piatimi miliónmi rokov. Pretože ľudia majú zložitý jazyk, zatiaľ čo humanoidné opice ho nemajú, biologické základy ľudského jazyka museli vzniknúť až po tomto oddelení.

Evolúcia nevytvorila schopnosť ľudskej reči úplne od počiatku v priebehu ostatných niekoľkých miliónov rokov, ale použila materiál, ktorý už bol k dispozícii, ktorý bol podstatne evolučne starší, ako práve vznikajúci jazyk[2]. Mnohé živočíšne druhy sú schopné sofistikovaných kognitívnych aktivít pri porozumení okolitého sveta a pri vzájomnej interakcii. Evolúcia využíva už existujúce štruktúry pre nové, často prekvapujúce ciele. Anatomická štruktúra mozgu humanoidných opíc je veľmi podobná anatomickej štruktúre ľudského mozgu, mnohé oblasti mozgu môžu byť navzájom medzi

---

[1] Fakulta informatiky a informačných technológií, Slovenská technická univerzita, 842 16 Bratislava,
E-mail: {pospichal, kvasnicka}@fiit.stuba.sk

[2] Podľa francúzskeho molekulárneho biológa F. Jacoba [23], darwinovská evolúcia organizmov so zložitou hierarchickou štruktúrou postupuje tak, že prebieha dominantne len na najvyšších hierarchických úrovniach, pričom nižšie hierarchické úrovne sú v podstate evolučne uzavreté – invariantné. Evolúcia môže zasahovať do nižších hierarchických úrovni len veľmi vzácne a aj to len formou tzv. „bricolage“ (fušerstva), t. j. používa štruktúry, ktoré sú na danej úrovni bezprostredne k dispozícii. Napríklad, biochémia bunky bola evolučne uzavretá už pred miliardami rokov, po vzniku mnohobunkových organizmov evolúcia len veľmi vzácne zasahuje na tejto nízkej hierarchickej úrovni (preto môže existovať univerzálne platný vedný odbor biochémia). V prípade fatálnej potreby sa evolúcia pokúša aj na tejto úrovni pomocou „bricolage“ riešiť vzniknutý problém. Takouto „bricolage“ môže byť aj evolučný vznik univerzálnej gramatiky, keď evolúcia použila „predpripravené“ štruktúry v mozgu, ktoré pôvodne vykonávali iné funkcie.

opicami a ľuďmi pomerne presne identifikované. Oblasti ľudského mozgu, ktoré sú označené ako centrum reči, u opíc sú identifikované ako oblasti pre kontrolu svalov tváre a pre analýzu zvukového vstupu [51]. Táto skutočnosť asi podstatne uľahčila evolučné „predrôtovanie“ týchto oblastí pre účely vzniku ľudskej reči.

Z týchto úvah vyplýva, že ľudský jazykový inštinkt nie je dielom „zázračnej“ mutácie, alebo povedané poetickým jazykom Richarda Dawkinsa, *náhla inšpirácia slepého hodinára riadiaceho evolúciu* [11], ale dôsledkom niekoľko sto miliónov rokov, v ktorých evolúcia experimentovala so živočíšnou kogníciou. Jazyk umožnil našim predchodcom výmenu ideí, skúseností a umožnil tiež riešiť problémy paralelným spôsobom. Adaptívny význam ľudského jazyka je nespochybniteľný. Spolupráca pri lovení, plánovanie, koordinovanie aktivít, podieľanie sa na spoločných úlohách, sociálne väzby, manipulácia alebo podvádzanie, toto všetko môže mať úžitok zo zvyšovania schopnosti komunikovať.

Prírodný výber zohral dôležitú úlohu pri evolúcii živočíšnej komunikácie. Jazykovedec Ray Jackendoff poukázal na skutočnosť, že v ľudskom jazyku prežívajú určité štruktúry, ktoré boli formované v praveku ľudskej existencie [21,22]. Taktiež identifikoval v gramatike moderných jazykov „fosílie“, ktoré boli súčasťou primitívnych komunikačných systémov. Tieto úvahy Jackendoffa sú založené na známej Bickertonovej myšlienke [4], že moderné jazyky sa vyvinuli z „prajazyka“, ktorého fosílne zbytky stále môžu byť nájdené v ľudskom mozgu. Prajazyk vzniká vždy tam, kde používanie rozvinutých jazykov je násilne prerušené; napr. vznik jazyka „pidgin“ alebo jazyk detí, ktoré boli ťažko deprivované nedostatkom sociálnej interakcie (najznámejší príklad je dievča Genie, ktorá bola slepá a hluchonemá). *Tieto skutočnosti môžu slúžiť ako biologické a lingvistické skutočnosti, ktoré podporujú ideu postupného evolučného vzniku jazyka*.

Ľudský jazyk je jedinečná biologická vlastnosť nášho druhu, ktorá zásadným spôsobom ovplyvnila správanie ľudí a ich výskyt na Zemi. Pochopenie evolučných mechanizmov emergencie ľudského jazyka patrí medzi najzaujímavejšie problémy súčasného darwinizmu [38], pričom sa študuje význam prírodného výberu pre emergenciu základných vlastností ľudského jazyka, akými sú znaky, slová, syntaktická komunikácia a gramatika [40]. Cieľom tejto práce je poskytnúť evolučný pohľad na počiatočné etapy vzniku jazyka, teda ako boli postupnosti foném spájané do slov, postupnosti slov do viet v rámci primitívnej gramatiky. Pritom bude skúmaný aj význam gramatiky pre ľudský jazyk. Ďalším štádiom výskumu by logicky bolo štúdium rozvinutej gramatiky, ako je napríklad otázka existencie Chomskeho univerzálnej gramatiky [9,10,39] pri akvizícii jazyka dieťaťom. Toto posledné, najpokročilejšie štádium gramatiky ale vyžaduje ďalšie predpoklady k úspešnému vývoju, ako sú napr. mémy [12,13,27] alebo existencia Baldwinovho efektu [1]. Ich význam pre akvizíciu jazyka bol už úspešne modelovaný tak pre mémy [27,31,32,33,53], čo je podporované aj prístupom k jazyku ako ku kultúrnemu fenoménu [5,8], ako aj pre Baldwinov efekt [2,6,17,34], tu sa nimi však nebudeme zaoberať. Jadrom nášho príspevku je modelovanie vzniku „prajazyka“ (text vznikol s použitím prác [35]).

Evolučné pohľady na jednotlivé aspekty jazyka sú v súčasnosti intenzívne študované teoretickým biológom Martinom Nowakom so spolupracovníkmi [24,25,41-49]. Štúdie tohto typu môžu byť chápané ako pokračovanie teoretických snáh Manfreda Eigena a Petra Schustera [14,15] pri rozvoji a aplikácii teórie replikátorov ako univerzálneho nástroja pre štúdium Darwinovej evolúcie nielen v biológii, ale aj v iných

vedných prírodovedných a humanitných oblastiach (vznik univerzálneho darwinizmu [12,13]).

2 Multiagentové systémy a univerzálny darwinizmus

Kombinácia multiagentových systémov [16,26] s univerzálnym darwinizmom [12,13] patrí v súčasnosti medzi progresívne prístupy umelej inteligencie [16] k štúdiu evolučnej emergencie kolektívnych a koordinovaných aktivít, ktoré pôvodne v systéme neboli prítomné, a ktoré sú výsledkom elementárnych interakcií medzi dvojicami agentov.

2.1 Multiagentový systém

Pre naše potreby bude *multiagentový systém* [16,26] špecifikovaný ako množina rovnakých (homogénnych) agentov

$$P = \{A_1, A_2, ..., A_r\}$$

ktoré majú predpísanú intencionalitu vykonávať nejakú aktivitu (napr. spolupracovať, komunikovať, pohybovať sa v prostredí a vyhýbať sa prekážkam, zbierať potravu,...). Táto všeobecná špecifikácia multiagentových systémov bude plne postačujúca pre naše ďalšie účely, nie je potrebné špecifikovať ich implementáciu, komunikačné protokoly a pod.

2.2 Darwinovské systémy a univerzálny darwinizmus

V súčasnosti je paradigma darwinovskej evolúcie široko aplikovaná v rôznych oblastiach ľudského poznania (poznáme evolučnú ekológiu, evolučnú ekonomiku, evolučnú psychológiu, evolučnú lingvistiku, evolučnú epistemológiu, evolučnú výpočtovú vedu, evolučnú fyziku...). Spoločným leitmotívom všetkých týchto nových vedných oblastí je použitie princípov darwinovskej evolučnej teórie, ktorá je v súčasnosti [13,38] charakterizovaná ako *univerzálny algoritmus* s platnosťou nielen v biológii, ale aj vo všetkých oblastiach ľudského poznania, kde sme schopní vyabstrahovať informačné entity – replikátory, ktoré majú schopnosť reprodukovať sa s variáciami, a medzi ktorými prebieha prírodný výber. Prvý, kto na túto univerzálnosť darwinovskej evolúcie upozornil, bol americký populačný genetik Sewall Wright. Ten v 30. rokoch minulého storočia charakterizoval *evolúciu ako optimalizáciu na povrchu fitnes funkcie* (fitness landscape), kde sa hľadá genotyp odpovedajúci globálnemu maximu. Z tohto pohľadu môžeme darwinovskú evolúciu charakterizovať ako univerzálnu paradigmu schopnú riešiť komplikované optimalizačné problémy vyskytujúce sa v širokom spektre vedných oblastí, od prírodných vied až po spoločenské vedy. V informatike našla táto zaujímavá idea svoj odraz už pred viac ako 30 rokmi, keď John Holland [19,28] vynašiel genetické algoritmy, ktoré je možno chápať ako algoritmizáciu darwinovskej evolúcie, a ktoré sa stali v súčasnosti búrlivo sa rozvíjajúcou oblasťou informatiky. Ich použitie umožňuje riešiť mnohé problémy, ktoré sa donedávna zdali len ťažko (ak vôbec) riešiteľnými. V umelej inteligencii použitie genetických algoritmov otvára úplne nové možnosti pre tvorbu stratégií a modelov (ktoré reprezentujú obvykle veľmi zložité problémy pre človeka), ich tvorba je prenechaná počítaču pomocou prístupu darwinovskej evolúcie „in silico“. Najvšeobecnejšia formulácia základnej idey univerzálneho darwinizmu je uskutočnená

pomocou koncepcie ***darwinovského systému*** založeného na nasledujúcich dvoch postulátoch:
(1) *Darwinovský systém sa skladá z populácie replikátorov* – jedincov/objektov, ktoré sú za určitých vhodných podmienok schopné *replikácie* – rozmnožovania. Replikačný proces spočíva v „kopírovaní" jedincov do populácie, pričom toto „kopírovanie" sa uskutočňuje s *určitými malými chybami*. Z tohto predpokladu nepresnej replikácie (s malými chybami) vyplýva, že *populácia nie je homogénna*, obsahuje jedincov – replikátory, ktoré vykazujú určitú malú variabilitu.
(2) Každý replikátor populácie je ohodnotený *fitnes*; hodnotenie vyjadruje schopnosť replikátora prežiť a úspešne vstupovať do replikačného procesu. *Prírodný výber* v darwinovskom systéme spočíva v tom, že jedinci populácie nie sú vyberaní do replikačného procesu náhodne, ale s *pravdepodobnosťou úmernou ich fitnes* (hovoríme, že výber sa deje kvázináhodne).

V definícii darwinovského systému sa vyskytuje niekoľko pojmov, ktoré musia byť bližšie objasnené. Pojem „replikátor" je potrebné chápať ako relatívne samostatnú entitu, ktorá existuje v nejakom prostredí a vytvára spolu s ostatnými replikátormi populáciu. Najjednoduchší prístup k implementácii replikátora je jeho stotožnenie s nejakou informáciou, ktorá kóduje jeho „telo". Z tohto dôvodu môžeme využívať biologickú terminológiu a rozlišovať dve rôzne špecifikácie replikátora, jeho *genotyp* a *fenotyp*. Genotyp je informácia o architektúre replikátora, zatiaľ čo fenotyp je „telo" replikátora. Vzťah medzi genotypom a fenotypom nebudeme bližšie špecifikovať, obvykle je silne závislý na type darwinovského systému. Čo bude pre nás dôležité, je skutočnosť, že fitnes replikátora je určený schopnosťou fenotypu prežívať v danom prostredí a vstupovať do výhodných interakcií s inými replikátormi z populácie. Musíme však podotknúť, že môžu existovať darwinovské systémy, kde odlíšenie fenotypu od genotypu neplatí, kde sa genotyp kopíruje – replikuje iným zariadením, ako vlastným fenotypom. Dobrým príkladom tejto situácie sú biologické a počítačové vírusy, ktoré k vlastnej replikácii využívajú systémy, v ktorých parazitujú.

Predpokladajme, že replikátor je reprezentovaný svojím *genotypom* $\boldsymbol{x}$, ktorý, ako už bolo poznamenané, obsahuje úplnú informáciu o architektúre a vlastnostiach replikátora. *Populácia* replikátorov je multimnožina genotypov

$$P = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_p\}$$

Vzájomný vzťah medzi genotypom a fitnes v univerzálnom darwinizme môže byť realizovaný dvoma rozličnými spôsobmi. V *prvom prístupe* je fitnes genotypu určené priamo genotypom, čiže fitnes je určené zobrazením $G \xrightarrow{f} [0,\infty)$, alebo

$$fitnes(\boldsymbol{x}) = f(\boldsymbol{x})$$

Tento prístup sa využíva vtedy, keď fitnes replikátorov je určený zložením ich genotypov a nie vzájomnou interakciou, ktorá by predstavovala selekčnú výhodu. V *druhom prístupe* je fitnes replikátorov určený vzájomnou párovou interakciou replikátorov z populácie,

$$fitnes(\boldsymbol{x}) = f(\boldsymbol{x}) = \sum_{\substack{\boldsymbol{x}' \in P \\ (\boldsymbol{x}' \neq \boldsymbol{x})}} F(\boldsymbol{x}, \boldsymbol{x}')$$

kde $F(\boldsymbol{x}, \boldsymbol{x}')$ vyjadruje párový príspevok k fitnes replikátora $\boldsymbol{x}$ ako dôsledok interakcie s iným replikátorom $\boldsymbol{x}'$.

Replikačný proces v univerzálnom darwinizme sa zvykne rozlišovať ako unárny (asexuálny, v našich ďalších úvahách budeme uvažovať hlavne tento spôsob replikácie) a binárny (sexuálny). Pri unárnej replikácii sa jeden replikátor – rodič zúčastňuje procesu, zatiaľ čo pri binárnej replikácii sa zúčastňujú procesu dva replikátory – rodičia. Rodičia (rodič) sú kvázináhodne vybraní z populácie v závislosti na ich fitnes (replikátory s väčším fitnes vstupujú do replikácie s väčšou pravdepodobnosťou) a produkujú nové replikátory – potomkov. Budeme rozlišovať tieto tri zložky replikačného procesu

(1) *selekcia* rodiča,
(2) *replikácia* rodiča, pričom vzniká potomok, a
(3) *návrat* potomka do populácie.

V prvom kroku replikácie sa vyberie pomocou stochastického operátora O_{select} z populácie P jeden replikát $\boldsymbol{x}_{parent} = O_{select,fitness}(P)$ tak, že pravdepodobnosť jeho výberu je úmerná fitnes. V druhom kroku použitím stochastického operátora replikácie O_{repli} z rodiča dostaneme potomka $\boldsymbol{x}_{offspring}$, $\boldsymbol{x}_{offspring} = O_{repli}(\boldsymbol{x}_{parent})$. Poznamenajme, že tento replikačný operátor nespočíva len v jednoduchom kopírovaní rodiča na potomka, ale taktiež v tom, že obsahuje náhodné mutácie, ktoré menia pôvodný kód rodiča. V treťom kroku sa rieši problém návratu potomka do populácie, v jednoduchej verzii tento krok reprodukčného procesu budeme riešiť tak, že náhodne vybraný replikátor $\boldsymbol{x}' = O_{select}(P)$ je nahradený v populácii P potomkom,

Algoritmus X.1.

```
P:=náhodne vygenerovaná populácia replikátorov;
t:=0;
pokiaľ t<tmax urob
    t:=t+1;
    xparent:=Oselect,fitness(P);
    xoffspring:=Orepli(xparent);
    x':=Oselect(P);
    P:=(P+xoffspring)-x';
```

Algoritmus je inicializovaný náhodnou generáciou populácie. Evolúcia populácie je obsiahnutá vo vnútri **pokiaľ-urob** cyklu, ktorý sa opakuje $t_{\max}$-krát. V rámci jednej epochy sa vytvorí potomok, ktorý v populácii nahradí náhodne vybraný replikátor.

Hybnou silou evolúcie darwinovského systému je (1) prírodný výber a (2) spontánne mutácie. Z tohto pohľadu môžeme povedať, že evolúcia systému prebieha na hrane chaosu a poriadku. Kvázideterministické mechanizmy selekcie najlepšie prispôsobených replikátorov pomocou prírodného výberu a neustále sa meniaca populácia replikátorov dôsledkom stochastických mutácií pri replikačnom procese, sú hlavným zdrojom evolúcie systému, postupného zvyšovania priemerného fitnes v celej populácii.

3 Eigenove replikátory

Manfred Eigen publikoval počiatkom 70. rokov dôležitý článok „*Self organization of matter and the evolution of biological macro molecules*“ [14,15], kde postuloval hypotetický chemický systém obsahujúci biomakromolekuly nazvané *replikátory*. Tento chemický systém napodobňuje Darwinovu evolúciu už na abiotickej úrovni. Eigen spolu so Schusterom diskutovali tento modelový chemický systém ako potenciálne možný abiotický mechanizmus hnacej sily pre zvyšovanie zložitosti na hraniciach medzi biotickými a abiotickými systémami.

Uvažujme *replikátory* $X_1, X_2, ..., X_n$, ktoré sú schopné nasledujúcich chemických reakcií:

$$X_i \xrightarrow{k_i} X_i + X_i \quad (i = 1, 2, ..., n)$$

$$X_i \xrightarrow{\phi} \varnothing \quad (i = 1, 2, ..., n)$$

Prvá reakcia znamená, že molekula X_i sa replikuje s rýchlostnou konštantou k_i, druhá reakcia znamená zánik (extinkciu) replikátora X_i s rýchlostnou konštantou ϕ (tento parameter sa nazýva „*zrieďovací tok*“ a bude špecifikovaný neskoršie). Použitím zákonitostí chemickej kinetiky získame systém diferenciálnych rovníc, ktorý popisuje dynamické vlastnosti replikátorového systému

$$\dot{x}_i = x_i (k_i - \phi) \quad (i = 1, 2, ..., n)$$

kde $0 \le x_i \le 1$ je koncentrácia (frekvencia výskytu) replikátora X_i v chemostate (populácii) P. Zrieďovací tok ϕ je voľný parameter, ktorý je určený takým spôsobom, že je splnená nasledujúca podmienka: súčet časových derivácií koncentrácií $\dot{x}_i$ je nulový, $\sum \dot{x}_i = 0$, potom

$$\dot{x}_i = x_i \left(k_i - \sum_{j=1}^{n} k_j x_j \right) \quad (i = 1, 2, ..., n)$$

kde podmienka $\sum x_i = 1$ bola použitá bez straty všeobecnosti našich úvah. Analytické riešenie tohto systému má tvar

$$x_i(t) = \frac{x_i(0) e^{k_i t}}{\sum_{j=1}^{n} x_j(0) e^{k_j t}}$$

Toto riešenie má asymptotickú vlastnosť, podľa ktorej len jeden typ replikátora v systéme prežíva (s maximálnou rýchlostnou konštantou k_{max}), ostatné replikátory zanikajú

$$\lim_{t \to \infty} x_i(t) = \begin{cases} 1 & (\text{pre } k_i = k_{max} = max\{k_1, ..., k_n\}) \\ 0 & (\text{inak}) \end{cases}$$

Obrazne povedané, každý typ molekuly – replikátora môže byť chápaný ako biologický druh s fitnes špecifikovaným rýchlostnou konštantou k. V chemickom reaktore (tiež nazývanom *chemostat*) „prežívajú“ len tie replikátory – druhy, ktoré sú najlepšie prispôsobené, t. j. majú najväčšiu rýchlostnú konštantu k_{max}, všetky ostatné molekuly s menšími rýchlostnými konštantami zanikajú. Podmienka konštantnosti sumy koncentrácií (t. j. $\sum x_i = 1$) zavádza „selekčný tlak“ medzi replikujúcimi sa molekulami,

len tie molekuly prežijú, ktoré sú najlepšie prispôsobené s maximálnou rýchlostnou konštantou. Zrieďovací tok ϕ potom môžeme interpretovať ako priemerný fitnes celej populácie replikátorov.

Vyššie popísaný jednoduchý systém replikátorov môže byť jednoducho zovšeobecnený tak, že replikácia je doprevádzaná aj mutáciami, potom systém diferenciálnych rovníc má tvar

$$\dot{x}_i = x_i\left(k_{ii} - \phi\right) + \sum_{\substack{j=1 \\ (j \neq i)}}^{n} k_{ji} x_j \quad \left(i = 1, 2, \ldots, n\right)$$

kde k_{ij} sú rýchlostné konštanty reakcie s mutáciou

$$X_i \xrightarrow{k_{ij}} X_i + X_j \quad \left(i, j = 1, 2, \ldots, n\right)$$

Budeme predpokladať, že matica rýchlostných konštánt $\boldsymbol{K} = \left(k_{ij}\right)$ má dominantné diagonálne elementy, t. j. nediagonálne elementy sú omnoho menšie ako diagonálne elementy. Tento predpoklad je založený na postuláte (odpozorovanom z prírody), že replikácia s mutáciou je veľmi vzácny proces, nemôže prekročiť určitú malú prahovú hodnotu, v opačnom prípade by zanikla vlastnosť systému, že víťazia najlepšie prispôsobené replikátory (spolu so svojimi blízkymi mutantami). Zrieďovací tok ϕ je určený podmienkou nulovosti sumy časových derivácií koncentrácií $\sum \dot{x}_i = 0$

$$\phi = \sum_{i,j=1}^{n} k_{ji} x_j$$

Analytické riešenie systému diferenciálnych rovníc vyzerá podobne ako pre systém s presnou replikáciou bez mutácií, v tomto prípade je však asymptotické riešenie doprevádzané aj svojimi blízkymi mutantami. To znamená, že zavedenie mutácií do systému replikátorov nemení dramaticky všeobecné vlastnosti systému bez mutácií. V tomto všeobecnejšom prípade, keď uvažujeme taktiež existenciu pomerne vzácnych mutácií pri replikačnom procese, taktiež prežívajú (pre $t \to \infty$) najlepšie prispôsobené replikátory s maximálnou rýchlostnou konštantou k_{max}. Tieto víťazné molekuly sú však v tomto všeobecnejšom prípade doprevádzané aj suitou svojich blízkych mutantov.

Ako už isto pozorný čitateľ zistil, naznačená Eigenova teória replikátorov je založená na predpoklade, že fitnes replikátorov (rýchlostné konštanty) je plne určený len štruktúrou samotných replikátorov a nie ich vzájomnou interakciou v populácii. Ako už bolo spomenuté pri prezentácii univerzálneho darwinizmu, existuje aj druhý pohľad na formuláciu teórie replikátorov, kde ich fitnes je určené práve ich vzájomnou interakciou v rámci populácie. To znamená, že pri tomto druhom prístupe sa ohodnotenie replikátorov fitnes dynamicky mení v závislosti na zložení populácie. Nech „chemická“ reakcia špecifikujúca replikátorový systém s mutáciami má tento tvar

$$X_i \xrightarrow{q_{ij} f_i(\boldsymbol{x})} X_i + X_j \quad \left(i, j = 1, 2, \ldots, n\right)$$

To znamená, že rýchlosť tejto reakcie je určená nielen „rýchlostnou konštantou“ q_{ij}, ale aj fitnes samotného replikátora X_i. Potom diferenciálne rovnice popisujúce dynamiku tohto systému majú tvar

$$\dot{x}_i = x_i\left(q_{ii}f_i(\boldsymbol{x}) - \phi\right) + \sum_{\substack{j=1\\(j\neq i)}}^{n} q_{ji}f_j(\boldsymbol{x})x_j \quad (i = 1,2,\ldots,n)$$

kde zrieďovací člen je určený podmienkou nulovosti súčtu časových derivácií, $\sum \dot{x}_i = 0$,

$$\phi = \sum_{i,j=1}^{n} q_{ji}f_j(\boldsymbol{x})x_j = \sum_{j=1}^{n}\left(\sum_{i=1}^{n} q_{ji}\right)f_j(\boldsymbol{x})x_j \approx \sum_{j=1}^{n} q_{jj}f_j(\boldsymbol{x})x_j$$

kde na pravej strane je uvedený približný tvar zrieďovacieho člena, ktorý vznikne zanedbaním nediagonálnych príspevkov q_{ij}. Táto približná formula pre zrieďovací člen sa stáva presnou formulou, keď rýchlostné konštanty q_{ij} budeme interpretovať ako pravdepodobnosti „premeny“ replikátora X_i na replikátor X_j, potom $\sum_i q_{ji} = 1$. Teória týchto replikátorových diferenciálnych rovníc s explicitným zahrnutím fitnes jednotlivých replikátorov je študovaná Hofbauerom a Sigmundom v známej knihe „*Evolutionary Games and Replicator Dynamics*“ [18], ich modelovanie evolúciou je diskutované napr. v [30].

Ako budeme používať výsledky uvedené v tejto úvodnej časti našej publikácie? Použitý prístup, ako už bolo uvedené, je zmesou multiagentového prístupu a univerzálneho darwinizmu v Eigenovej replikátorovej verzii. Budeme predpokladať, že máme populáciu agentov a budeme u nich sledovať evolučnú emergenciu vybranej jazykovej schopnosti. Táto možnosť kombinácie multiagentových systémov s evolúciou vyplýva z nášho implicitného predpokladu, že schopnosť agentov komunikovať medzi sebou zvyšuje fitnes. Ináč povedané, schopnosť agentov efektívne komunikovať podstatne zvyšuje ich selekčnú výhodnosť pred inými agentmi, ktoré túto schopnosť nemajú. Z tejto skutočnosti taktiež vyplýva, že fitnes agentov je určený pomocou druhého postupu z univerzálneho darwinizmu, kde fitnes jedného agenta závisí na schopnosti efektívne komunikovať s ostatnými agentami z celej populácie v rámci zvoleného komunikačného – jazykového systému. Potom platí

$$f(A) = \frac{1}{|P|-1}\sum_{\substack{A'\in P\\(A'\neq A)}} F(A,A')$$

kde $F(A,A')$ reprezentuje „platbu“, ktorú získa agent A pri komunikácii s iným agentom A', veľkosť tejto platby je úmerne závislá na úspešnosti tejto komunikácie. Tento druhý prístup k formulácii replikátorových rovníc aj napriek svojej biologickej plauzibílnosti má jedno vážne ohraničenie, neexistuje ich teória popisujúca asymptotické vlastnosti riešení systému diferenciálnych rovníc, ktorá by jednoduchým spôsobom predpovedala existenciu asymptotických stavov ako tomu bolo pre replikátorové rovnice s nemennými rychlostnými konštantami. Tento nedostatok je čiastočne riešený teóriou evolučne stabilných riešení, ktorá bola vypracovaná evolučnými biológmi Maynard-Smithom a Riceom [36,37] pred 30 rokmi, keď použili teóriu hier (ktorá bola vtedy známa hlavne v ekonómii) na štúdium konfliktov v živočíšnej ríši.

4 Komunikácia pomocou ľubovoľných signálov

V tomto odseku budeme študovať evolučnú emergenciu najjednoduchšieho signalizačného systému, v ktorom sú ľubovoľné zvukové signály priradené určitým špecifickým situáciám – objektom [7,20,24,25,29,41-49]. Uvažujme populáciu agentov

$$P = \{A_1, A_2, \ldots, A_r\}$$

ktorí medzi sebou po dvojiciach komunikujú pomocou elementárneho systému, kde k jednotlivým objektom sú priradené špecifické zvuky (napr. fonémy). Nech každý agent je špecifikovaný asociačnou maticou $\boldsymbol{A} = \left(a_{ij}\right)_{j=1,2,\ldots,n}^{i=1,2,\ldots,n}$, kde n je tak počet objektov, ako aj zvukov. Maticový element $a_{ij} \geq 0$ špecifikuje silu asociácie medzi i-tym objektom a j-tym zvukom. Študujme elementárny komunikačný akt medzi agentmi A a A', ktorí majú asociačné matice $\boldsymbol{A}$ resp. $\boldsymbol{A'}$. Agent A (deklarovaný ako *hovorca*) si vytvorí zo svojej asociačnej matice novú maticu $\boldsymbol{P} = \left(p_{ij}\right)_{j=1,2,\ldots,n}^{i=1,2,\ldots,n}$, ktorá určuje pravdepodobnosť toho, že agent A interpretuje i-ty objekt j-tym zvukom, potom

$$p_{ij} = \frac{1}{\sum_{k=1}^{n} a_{ik}} a_{ij}$$

Analogicky, agent A' (deklarovaný ako poslucháč) vytvorí zo svojej asociačnej matice $\boldsymbol{A'}$ novú maticu $\boldsymbol{Q'} = \left(q'_{ji}\right)_{i=1,2,\ldots,n}^{j=1,2,\ldots,n}$, ktorej maticové elementy sú pravdepodobnosti toho, že j-ty zvuk je interpretovaný i-tym objektom.

$$q'_{ji} = \frac{1}{\sum_{k=1}^{n} a'_{kj}} a'_{ij}$$

Pri vzájomnom orientovanom komunikačnom akte medzi agentmi A a A' agent–hovorca A posiela agentovi – poslucháčovi A' signál kódujúci nejaký objekt. Úspešnosť tejto komunikácie je ohodnotená platbou $F(A,A')$, ktorá je určená vzťahom

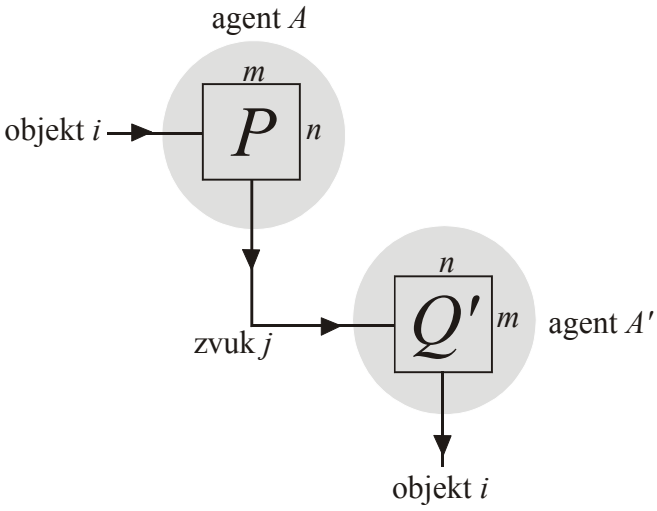
$$F\left(A, A'\right) = \sum_{i=1}^{n} \sum_{j=1}^{n} p_{ij} q'_{ji} = Tr\left(PQ'\right)$$

Takto definovaná platba pre komunikáciu $A \to A'$ obsahuje všetky možné objekty interpretované všetkými možnými zvukmi, pozri obrázok 1.

Každý agent $A \in P$ je ohodnotený *fitnes*

$$f\left(A\right) = \frac{1}{|P|-1} \sum_{\substack{A' \in P \\ (A' \neq A)}} F\left(A, A'\right)$$

Fitnes agenta A odráža jeho úspešnosť v komunikácii ako hovorcu, čím je jeho fitnes väčšie, tým úspešnejšie komunikoval s ostatnými agentmi (poslucháčmi) z populácie.

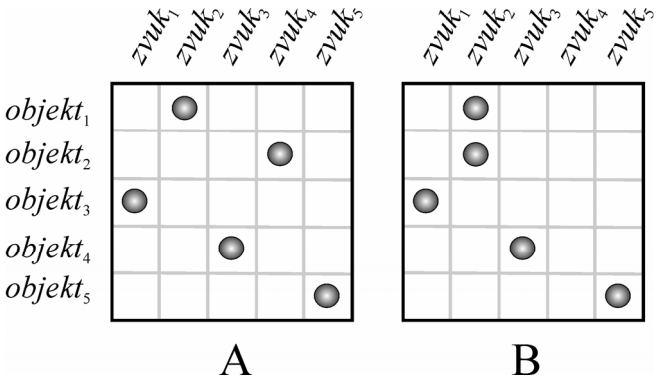


Obrázok 1. Schéma komunikácie $A \to A'$, platba tejto komunikácie $F(A,A')$ je určená ako súčet súčinov pravdepodobností $p_{ij}q'_{ji}$ pre všetky možné objekty a zvuky.

Cieľom evolúcie populácie agentov je nájsť také optimálne riešenie pre asociačnú maticu $\boldsymbol{A}$, ktoré maximalizuje fitnes všetkých agentov populácie

$$\left(\boldsymbol{A}\right)_{opt} = arg\ \max_{\tilde{\boldsymbol{A}}} f\left(\tilde{\boldsymbol{A}}\right)$$

Aký tvar má toto optimálne riešenie? Túto úlohu vyriešime tak, že budeme hľadať za akých podmienok je platba medzi dvojicou vybraných agentov $A \to A'$ maximálna, $F(A,A')=max$.



Obrázok 2. Znázornenie rôznych typov asociačných matíc pre n = 5 (t. j. 5 objektov je klasifikovaných pomocou 5 zvukov), vyhovujúcich podmienke, že každý riadok obsahuje práve jeden jednotkový prvok. (A) Asociačná matica popisuje jednoznačné priradenie medzi objektmi a zvukmi, je charakterizovaná ako permutačná matica, platba sa rovná 5. (B) Asociačná matica reprezentuje prvý a druhý objekt rovnakým zvukom, tento prípad odpovedá nejednoznačnému priradeniu medzi objektmi a zvukmi, platba je menšia a rovná sa 4.

Predpokladajme, že matica $\boldsymbol{Q}'$ agenta A' je zafixovaná, budeme hľadať optimálnu maticu $\boldsymbol{P}$ agenta A, ktorá maximalizuje platbu $F(A,A')$. Jednoduchými úvahami sa dá ukázať, že platba $F\left(A,A'\right) = Tr\left(\boldsymbol{PQ}'\right)$ nadobúda maximálnu hodnotu vtedy, keď matica $\boldsymbol{P}$

je vždy binárna matica, kde $p_{ij} = 1$ vždy tam, kde element q'_{ji} nadobúda maximálnu hodnotu pre *i*-ty stĺpec; všetky ostatné elementy matice ***Q'*** sú nulové. Týmto spôsobom sme dostali, že pre zafixovanú maticu ***Q'*** platba *F*(*A*,*A'*) je maximálna pre binárnu maticu ***P***. V druhom kroku budeme riešiť podobnú úlohu, avšak už budeme predpokladať, že matica **P** je binárna. Podobnými úvahami ako v predchádzajúcom kroku dostaneme, že matica ***Q'*** musí byť taktiež binárna, pričom platí $\boldsymbol{Q}' = \boldsymbol{P}^T$. Pomocou tohto výsledku sme taktiež dokázali, že maximálna platba pre komunikáciu $A \to A'$ je $F\left(A, A'\right) = n$, pozri obr. 2.

4.1 Replikácia – učenie agentov

Evolučný algoritmus obsahuje ako integrálnu časť replikáciu otcovského agenta, pričom vzniká agent potomok

$$A_{offspring} = O_{repli}\left(A_{parent}\right)$$

Pri tomto replikačnom procese sa vlastne tvorí z otcovskej asociačnej matice nová asociačná matica potomka, takže replikačný proces môžeme formálne prepísať do tvaru, kde explicitne vystupujú asociačné matice

$$\boldsymbol{A}_{offspring} = O_{repli}\left(\boldsymbol{A}_{parent}\right)$$

V tejto súvislosti musíme však poznamenať, že asociačná matica nepatrí do genotypu agenta, ale je len určitou „kultúrnou“ výbavou agenta, ktorej prenos z rodičov na potomkov sa zabezpečuje procesom učenia

$$\boldsymbol{A}_{offspring} = O_{learning}\left(\boldsymbol{A}_{parent}\right)$$

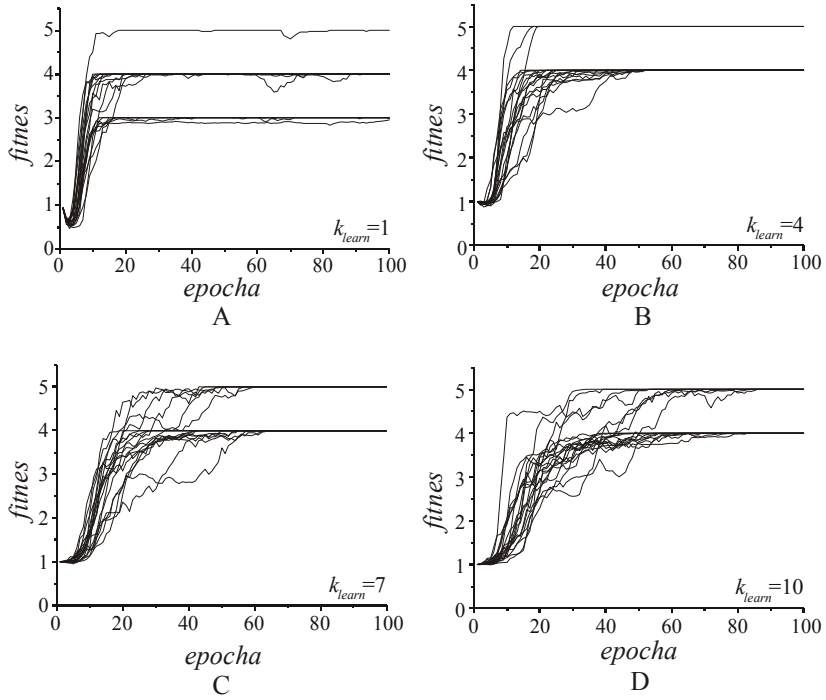
Preto súčasťou evolučného algoritmu musí byť aj špecifikácia procesu učenia, pri ktorom vzniká z rodičovskej asociačnej matice nová asociačná matica potomka. Z rodičovskej asociačnej matice vytvoríme maticu $\boldsymbol{P}_{parent}$ pomocou normalizácie riadkov príslušnej asociačnej matice $\boldsymbol{A}_{parent}$. V nasledujúcom kroku učenia vytvoríme asociačnú maticu potomka $\boldsymbol{A}_{offspring}$ tak, že jej jednotlivé maticové elementy opakovane (k_{learn} krát) zväčšujeme o jednotku s pravdepodobnosťou úmernou veľkosti príslušného elementu matice $\boldsymbol{P}_{parent}$ (vzorkovanie – sampling tejto matice). Tento proces vytvárania nových asociačných matíc schematicky znázorníme takto

$$\begin{array}{l}
\boldsymbol{A}^{(0)}_{parent} \to \boldsymbol{P}^{(0)}_{parent} \to \boldsymbol{A}^{(1)}_{offspring} \\
\qquad\qquad\qquad\qquad\qquad\quad \downarrow \\
\qquad\qquad\qquad\qquad\quad \boldsymbol{A}^{(1)}_{parent} \to \boldsymbol{P}^{(1)}_{parent} \to \boldsymbol{A}^{(1)}_{offspring} \\
\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \downarrow \\
\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \ldots\ldots\ldots\ldots
\end{array}$$

Dôležitým parametrom, ktorý špecifikuje tento proces učenia (t. j. vytváranie nových asociačných matíc potomkov a ***P*** matíc rodičov) je konštanta k_{learn}, ktorá určuje počet opakovaní elementárneho aktu učenia pre daný maticový element.

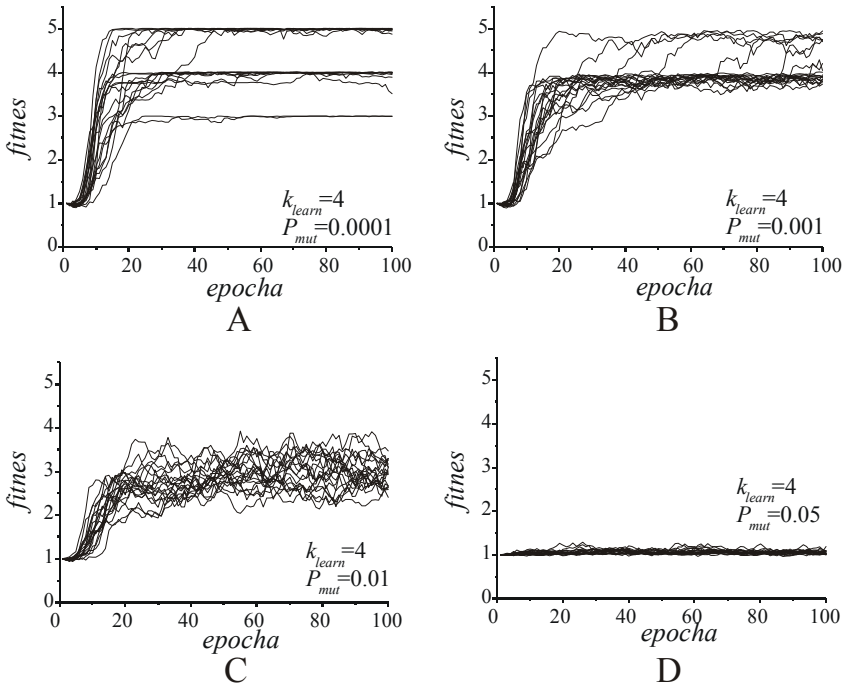
Proces učenia môže byť stochasticky porušený tak, že pri tvorbe asociačnej matice $\boldsymbol{A}_{offspring}$ pomocou vzorkovania matice $\boldsymbol{P}_{parent}$ s malou pravdepodobnosťou P_{mut} zavedieme poruchu. To znamená, že veľkosť maticových elementov vytváranej asociačnej matice $\boldsymbol{A}_{offspring}$ už nezávisí len od veľkosti jednotlivých elementov matice $\boldsymbol{P}_{parent}$, ale taktiež aj od náhodného procesu (v disjunkcii) špecifikovaného pravdepodobnosťou P_{mut}.

Získané numerické výsledky sú zhrnuté v obrázkoch 3 a 4. Diagramy z obrázku 3 obsahujú 20 nezávislých priebehov priemerného fitnes pre rôzne konštanty k_{learn}. Z týchto výsledkov vyplýva, že zvyšovanie veľkosti k_{learn} po prekročení určitej kritickej hodnoty už podstatne neovplyvňuje získané výsledky, ktoré približne s rovnakou pravdepodobnosťou dosahujú konečné fitnes 4 alebo 5. Avšak pokiaľ do evolúcie zahrnieme náhodné chyby pri učení popísané pravdepodobnosťou $\boldsymbol{P}_{mut}$, výsledky sa dramaticky menia v závislosti na tejto pravdepodobnosti. Ak pravdepodobnosť presiahne kritickú hodnotu ležiacu medzi 0.001 a 0.01 evolúcia sa začne správať stochasticky a prestane monotónne konvergovať k finálnej hodnote fitnes (4 alebo 5). Môžeme teda konštatovať, že proces evolúcie komunikácie pomocou jednoduchého princípu založeného na asociačnej matici je neobyčajne citlivý na chyby pri učení, keď rodič učí svojho potomka asociačnú maticu. Ak táto chybovosť presiahne určitú relatívne malú kritickú hodnotu, emergencia komunikácie má veľký výskyt chýb (chybných asociácií medzi objektom a signálom), alebo úplne zanikne.

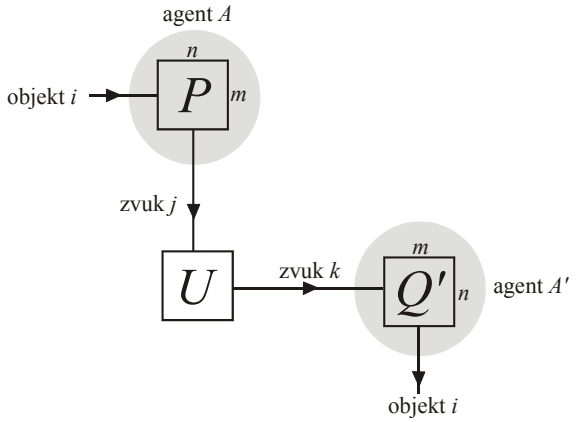


Obrázok 3. Znázornenie priebehov priemerného fitnes pre evolúciu asociačnej matice pre rôzne hodnoty počtu cyklov učenia k_{learn}, populácia obsahovala 100 agentov[3].

[3] Grafy z obrázkov 3 a 4 sú prevzaté z našej publikácie „*Replicator Theory of Coevolution of Chomsky's Universal Grammars and Generative Grammars*“, prezentovanej na konferencii SOFSEM 2004 [33].



Obrázok 4. Znázornenie priebehov priemerného fitnes pre evolúciu asociačnej matice, pričom proces učenia obsahuje náhodné poruchy špecifikované pravdepodobnosťou P_{mut}, populácia obsahovala 100 agentov a proces učenia bol špecifikovaný konštantou $k_{learn} = 4$.



Obrázok 5. Schéma komunikácie $A \rightarrow A'$ obsahujúcej šum pri prenose zvuku medzi agentami. Tento komunikačný šum je vyjadrený maticou $\boldsymbol{U}=(u_{ij})$, kde maticový element u_{ij} popisuje pravdepodobnosť chybného priradenia medzi i-tym a j-tym zvukom.

4.2 Kvalitatívna teória komunikácie s chybami

V predchádzajúcej časti tejto kapitoly bolo ukázané, že nepresnosti v rámci učenia majú silný dopad na emergenciu komunikácie v populácii agentov. Pokúsme sa tento dôležitý výsledok odvodiť v rámci kvalitatívneho modelu komunikácie v multiagentovom systéme. Štandardná komunikácia medzi dvojicou agentov $A, A' \in P$ je rozšírená o „šum“ vo zvukovom komunikačnom kanáli, pozri obr. 5.

Platba $F(A,A')$ pri komunikácii $A \to A'$ je určená vzťahom

$$F(A,A') = \sum_{i,j,k=1}^{n} p_{ij} u_{jk} q'_{ki} = Tr(PUQ')$$

Elementy matice $\boldsymbol{U}$ budú špecifikované pomocou podobnosti medzi zvukmi,

$$u_{ij} = \frac{s_{ij}}{\sum_{k=1}^{n} s_{ik}}$$

kde s_{ij} je podobnosť medzi zvukom i a j, ktorá je určená takto

$$s_{ij} = \begin{cases} 1 & (i = j) \\ \varepsilon & (i \neq j) \end{cases}$$

To znamená, že rôzne zvuky majú rovnakú malú podobnosť určenú malým kladným číslom ε. Potom diagonálne elementy u_{ii} sú určené takto

$$u_{ii} = \frac{1}{1+(m-1)\varepsilon}$$

Maximálna hodnota platby $F(A,A')$ je určená tak, že oba agenti majú rovnakú asociačnú maticu $\boldsymbol{A}$, pričom priradené matice $\boldsymbol{P}$ a $\boldsymbol{Q}$ sú permutačné matice spĺňajúce podmienku $Q = P^T$

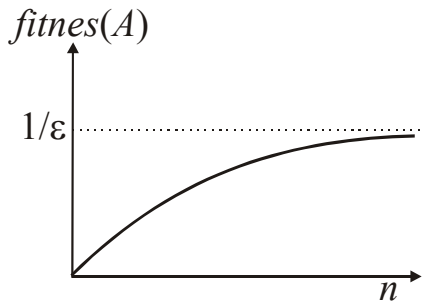
$$F(A,A') = \sum_{i,j,k=1}^{n} p_{ij} u_{jk} p_{ik} = \sum_{i=1}^{n} u_{ii} = \frac{n}{1+(n-1)\varepsilon}$$

Fitnes každého agenta A je určený ako suma všetkých možných platieb s inými agentmi z populácie

$$f(A) = \frac{1}{|P|} \sum_{A' \neq A} F(A,A') = \underbrace{\frac{|P|-1}{|P|}}_{\doteq 1} \frac{n}{1+(n-1)\varepsilon} \doteq \frac{n}{1+(n-1)\varepsilon}$$

Ak zväčšujeme počet objektov a zvukov, ktoré sú predmetom komunikácie, $n \to \infty$, potom fitnes v multiagentovom systéme nadobúda svoju maximálnu hodnotu (pozri obr. 6)

$$\lim_{n\to\infty} f(A) \doteq \lim_{n\to\infty} \frac{n}{1+(n-1)\varepsilon} = \frac{1}{\varepsilon}$$



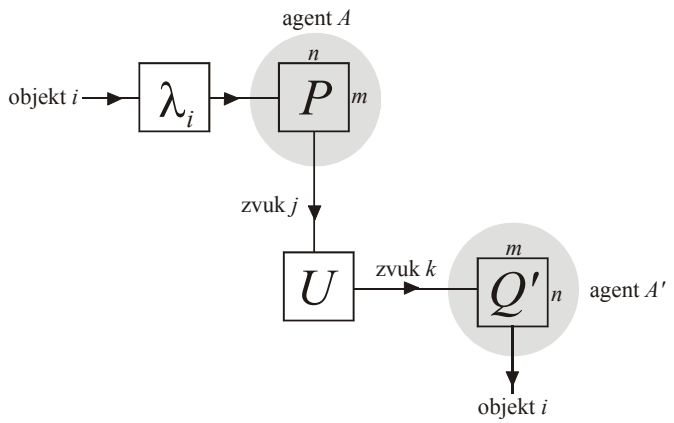
Obrázok 6. Priebeh maximálneho fitnes v populácii pri zvyšovaní počtu objektov a zvukov, ktoré sú predmetom komunikačných aktov medzi agentmi. Tento priebeh je monotónne rastúca a nasýtená funkcia dosahujúca maximálnu hodnotu $1/\varepsilon$.

Pomocou kvalitatívneho modelu sme ukázali, že fitnes je rastúcou funkciou počtu objektov/zvukov, pričom sa blíži k maximálnej hodnote $1/\varepsilon$ pre veľké hodnoty *n*. Ak komunikácia medzi agentmi neobsahuje chyby, potom maximálna hodnota fitnes je $f_{max} = n$. Avšak, ak komunikácia obsahuje chyby špecifikované parametrom ε, potom maximálna kapacita informačného prenosu je zhora ohraničená a rovná sa $1/\varepsilon$ zvukov. Pomocou týchto jednoduchých kvalitatívnych úvah sme ukázali, že ak sa pri komunikácii v multiagentovom systéme vyskytuje chyba, potom nie je možné efektívne zvyšovať fitnes agentov tým, že zvyšujeme počet zvukov; ako bolo ukázané, tento „priamočiary" prístup je nasýtený a v podstate je zhora ohraničený počtom zvukov $1/\varepsilon$.

Evolúcia komunikačného systému môže byť podstatne ovplyvnená tým, že objekty majú rôznu dôležitosť pre komunikujúcich agentov. Napríklad, leopard reprezentuje väčšie nebezpečenstvo ako obrovský had pytón, zvuk priradený objektu „leopard" musí mať väčšiu dôležitosť ako zvuk priradený objektu „pytón". Jednotlivé objekty sú ohodnotené pomocou veličín $\lambda_1 > \lambda_2 > ... > \lambda_n$, kde λ_i reprezentuje ohodnotenie dôležitosti *i*-teho objektu. Pre zjednodušenie našich nasledujúcich úvah predpokladajme, že objekty sú tak indexované, že postupnosť ich ohodnotení je klesajúca. Potom platba pre komunikáciu $A \to A'$ je určená vzťahom

$$F(A, A') = \sum_{i,j,k=1}^{n} \lambda_i p_{ij} u_{jk} q'_{ki} = Tr(\Lambda P U Q')$$

kde $\Lambda = diag(\lambda_1 > \lambda_2 > ... > \lambda_n)$ je diagonálna matica s diagonálnymi koeficientami špecifikovanými ako ohodnotenia objektov (pozri obr. 7) a λ_i sú náhodné čísla z otvoreného intervalu (0,1).



Obrázok 7. Rozšírenie schémy komunikácie z obr. 5 o ohodnotenie objektov.

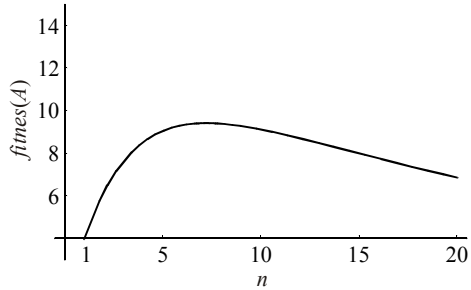
Maximálna hodnota platby sa získa pre takú populáciu agentov, ktoré majú rovnakú asociačnú maticu, pričom matica ***P*** je určená ako permutačná matica a ***Q'*** = $\boldsymbol{P}^T$, potom

$$F(A, A') = \sum_{i,j,k=1}^{n} \lambda_i p_{ij} u_{jk} p_{ik} = \sum_{i=1}^{n} \lambda_i u_{ii} = \frac{\sum_i \lambda_i}{1 + (m-1)\varepsilon}$$

Fitnes agenta A je určený vzťahom

$$f(A)=\frac{1}{|P|}\sum_{A'\neq A}F(A,A')=\underbrace{\frac{|P|-1}{|P|}}_{\doteq 1}\frac{\sum_{i=1}^{n}\lambda_i}{1+(n-1)\varepsilon}\doteq\frac{\sum_{i=1}^{n}\lambda_i}{1+(n-1)\varepsilon}$$

Priebeh fitnes agenta A pri raste parametra n už nie je monotónne rastúci (ako v predchádzajúcom prípade, keď všetky objekty mali rovnaké ohodnotenie), ale už vykazuje určité maximum, pozri obrázok 8. V tomto prípade prírodný výber ohraničuje počet používaných zvukov, a teda aj počet objektov, ktoré sú interpretované použitými zvukmi.



Obrázok 8. Priebeh maximálneho fitnes populácie agentov, ktorých komunikácia je založená na rovnakej asociačnej matici, pričom sa uvažuje šum a ohodnotenie objektov. Z grafu vyplýva, že fitnes rastie do určitej optimálnej hodnoty n, potom už len monotónne klesá.

4.3 Vznik slov

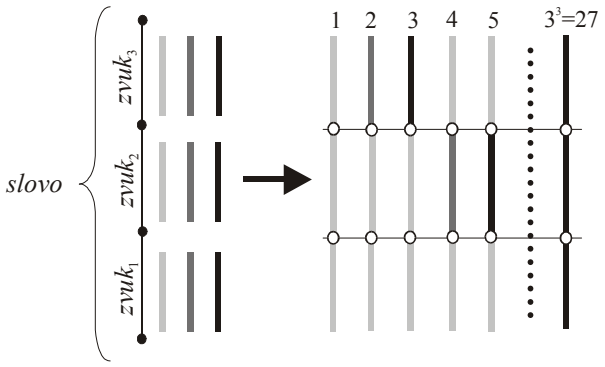
Asociačné matice sú vhodným prostriedkom k popisu nielen komunikácie medzi živočíchmi, kde špecifikujú asociácie medzi signálmi živočíchov a ich významom, ale slúžia aj k popisu lexikálnej matice ľudského jazyka. V prvom prípade je živočíšna komunikácia špecifikovaná asociáciami medzi zvukmi a ich významom (objektmi). V druhom prípade asociačná matica špecifikuje asociácie medzi tvarom slov a ich významom. Samozrejme, medzi týmito dvoma možnými prístupmi k použitiu asociačných matíc je potrebné si uvedomiť fundamentálnu rozdielnosť medzi živočíšnymi signálmi a slovnými formami. Živočíšna komunikácia je založená na pomerne obmedzenom repertoári 10-100 zvukov, zatiaľ čo ľudský jazyk používa niekoľko desaťtisíc slov. Naviac, ľudský jazyk využíva gramatiku k tvorbe rôznych kombinácií slov do viet, ktoré môžu mať diametrálne odlišný význam. Slová sú postupnosťou presne špecifikovaných stavebných blokov, nazývaných fonémy. Cieľom tejto podkapitoly bude poukázať na argumenty, prečo sú slová vytvárané z foném, ako určitého spôsobu k prekonaniu komunikačnej bariéry prístupu, ktorý používa len elementárne zvuky.

Pre jednoduchosť predpokladajme, že jazyk obsahuje m zvukov (foném), ktoré môžu byť kombinované do slov dĺžky (počtu foném) l, potom počet slov je $n = m^l$ (pozri obrázok 9). Maximálna platba medzi dvoma agentmi A a A' je určená vzťahom (pozri vzťahy z predchádzajúcej podkapitoly)

$$F(A,A')=\left(\frac{m}{1+(m-1)\varepsilon}\right)^{l}$$

pričom vyhovuje tejto asymptotickej podmienke

$$\lim_{m\to\infty} F\left(A, A'\right) = \left(\frac{1}{\varepsilon}\right)^{l}$$



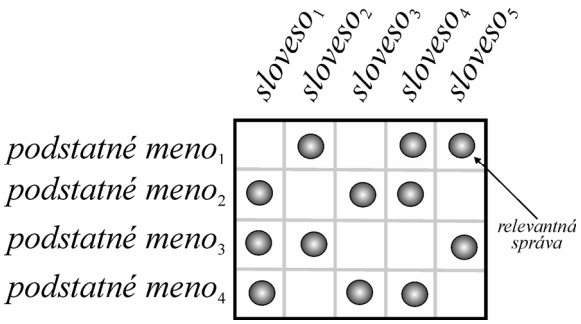
Obrázok 9. Znázornenie vzniku slov zo zvukov – foném, kombinatoriálnym spájaním vybraných foném vznikajú slova.

Z tohto vzťahu vyplýva, že ak necháme kombinovať zvuky do slov, komunikačná kapacita jazyka sa podstatne zvýši v závislosti na počtu foném (l) použitých pri kombinatoriálnej konštrukcii slov pomocou zvukov. Tento krok v jazykovej evolúcii môžeme pokladať za prechod od analógového k diskrétnemu systému, v ktorom komunikačné jednotky – slová už majú svoju štruktúru. Avšak musíme poznamenať, že ani tento prístup nerieši principiálne ohraničenie komunikačného systému; podobne ako aj v predchádzajúcich prípadoch, ak v tomto prístupe existuje evolučné ohraničenie komunikačnej kapacity, ďalšie zväčšovanie počtu slov zväčšovaním počtu foném už nezvyšuje fitnes živočícha. Fitnes sa stáva nasýteným, alebo dokonca začne mať klesajúci charakter (ak jednotlivým slovám pripíšeme rôzne hodnoty špecifikujúce ich dôležitosť). V existujúcich ľudských jazykoch len relatívne malý počet foném je používaný k tvorbe slov, čiže si evolúcia nezvolila tento prístup, kde sa komunikuje len prostredníctvom samotných slov (aj keď tvorených z foném), ale iný, založený na gramatike, ktorý dodal jazyku zhora neohraničenú komunikačnú silu.

4.4 Vznik elementárnej gramatiky

Jedna z najdôležitejších čŕt ľudského jazyka je gramatika, pomocou ktorej sa vytvárajú vety zo slov. *Gramatika* je formálny systém, ktorý zabezpečuje vzťah (reláciu) medzi vnútornou reprezentáciou významu, ktorý bude komunikovaný hovorcom, a tvarom vety reprezentujúcej tento vnútorný význam a poslanej poslucháčovi. Gramatika zabezpečuje ľudskému jazyku neohraničenú vyjadrovaciu schopnosť na základe vytvárania viet pomocou konečného lexikónu slov.

Ďalší možný krok v evolúcii jazyka po vzniku slov je teda vznik jednoduchých syntaktických štruktúr alebo gramatiky [24,25,46,47]. Gramatika spôsobuje dramatické zvýšenie informačného prenosu jazykom; jednoduché pomenovanie objektu signálom musí byť menej efektívne ako pomenovanie objektu a priradenie akcie, pozri obr. 10.



Obrázok 10. Jednoduchý model pre štúdium emergencie syntaktickej komunikácie, ktorá využíva signály zložené z dvoch častí, kde prvá časť špecifikuje objekt (podstatné meno), zatiaľ čo druhá časť špecifikuje akciu (sloveso). Možné signály môžeme vyjadriť v tvare tabuľky, kde riadky sú podstatné mená a stĺpce sú slovesá, prienik daného riadku a stĺpca odpovedá zloženine, ktorá príslušnému objektu priraďuje nejakú akciu (napr. leopard beží alebo opica spí).

Predpokladajme, že máme n objektov a h akcií, potom máme nh možných kombinácií, z ktorých však len určitá frakcia $N = \varphi nh$ je relevantných. Budeme rozlišovať dva rôzne prístupy k štúdiu tohto pohľadu na emergenciu elementárnej gramatiky.

Negramatický prístup, ktorý využíva $N = \varphi nh$ rôznych signálov pre identifikáciu relevantných signálov, maximálny fitnes tohto prístupu je

$$F_{ng} = \frac{N}{1 + (nh - 1)\xi}$$

kde ξ je malé kladné číslo špecifikujúce podobnosť medzi dvoma rôznymi zvukmi.
Gramatický prístup, ktorý využíva n slov pre objekty (podstatné mená) a h slov pre akcie (slovesá), maximálny fitnes tohto prístupu je

$$F_g = \frac{N}{\left[1 + (n-1)\xi\right]\left[1 + (h-1)\xi\right]}$$

To, či je evolučne výhodnejší gramatický alebo negramatický prístup, záleží na podmienke, či fitnes gramatického prístupu je väčší ako fitnes negramatického prístupu, $F_g > F_{ng}$, potom dostaneme

$$\frac{N}{\left[1 + (n-1)\xi\right]\left[1 + (h-1)\xi\right]} > \frac{N}{1 + (nh-1)\xi}$$

alebo

$$\xi < \xi_{max} = \frac{N - n - h + 1}{(n-1)(h-1)}$$

Z tejto podmienky môžeme odvodiť nutnú podmienku pre výhodnosť gramatického prístupu pred negramatickým prístupom. Pretože parameter ξ je kladný, potom čitateľ musí byť taktiež kladný

$$N - n - h + 1 \geq 1 \Rightarrow N \geq n + h$$

Gramatický prístup sa stáva výhodný vtedy, ak počet relevantných slov presiahne počet slov, z ktorých sa vytvárajú „zloženiny“ objekt – akcia. Môžeme teda konštatovať, že komunikačný systém, ktorý využíva zložené signály má väčšie potenciálne možnosti. Počet možných správ môže podstatne prevyšovať počet zložiek (slov), z ktorých sú tieto

správy zložené. Pre najjednoduchší komunikačný systém, kde každý význam je reprezentovaný iným signálom, každá správa musí byť naučená. Oproti tomu v komunikačnom systéme s elementárnou gramatikou, založenou na „zloženinách“ podstatné meno – sloveso, musia byť zapamätané len slová a nie celé vety, tieto sa vytvárajú ich jednoduchým spojením. Ak si uvedomíme túto skutočnosť, môžeme byť prekvapení, že živočíchy skoro vôbec nevyužívajú túto možnosť. Mozog nehumanoidných živočíchov zrejme ešte neobsahuje dostatočne rozvinuté štruktúry (kognitívne schopnosti), ktoré by tento efektívnejší komunikačný systém umožňovali (pozri obrázok 11).

"I'm afraid you misunderstood. ... I said I'd like a mango."

Obrázok 11. Vedecký vtip uverejnený v časopise *Nature*, ako súčasť krátkeho článku evolučného psychológa a kognitívneho vedca Stevena Pinkera [52], v ktorom vyjadril svoje kladné stanovisko k snahám študovať teoreticky alebo nasimulovať počítačom jednotlivé etapy evolúcie ľudského jazyka.

Evolučný model prechodu od elementárnej komunikácie ku komunikácii s elementárnou gramatikou musí vyhovovať dvom podmienkam, ktoré sú nutné k tomu, aby prírodný výber preferoval komunikáciu s elementárnou gramatikou: (1) Celkový počet relevantných správ musí prevyšovať určitú kritickú hodnotu. Len vtedy ak komunikačný systém dosiahol určitú kritickú veľkosť, sa používanie elementárnej gramatiky stáva selekčnou výhodou. Malé komunikačné systémy sú efektívne kódované pomocou elementárnych signálov používaných zvlášť pre každú správu. (2) Komunikačný systém s elementárnou gramatikou je schopný takého kódovania správ, kde sa jednotlivé zložky môžu vyskytovať v mnohých rôznych správach. V prípade, že sa každá zložka vyskytuje len v niekoľkých správach, potom existuje len malá šanca, aby takýto komunikačný systém konkuroval jednoduchému komunikačnému systému bez gramatiky.

Na základe týchto výsledkov môžeme konštatovať, že počet rôznych zvukov v prajazyku [3,4] (a teda aj počet objektov, ktoré sú zvukmi identifikované) je ohraničený. Dodanie nového zvuku zvýši počet identifikovaných objektov, avšak za cenu zvýšenia pravdepodobnosti chybnej komunikácie, celková schopnosť prenosu informácie nie je teda zvýšená. To znamená, že existuje vnútorná bariéra pre zvyšovanie počtu zvukov používaných na identifikáciu objektov, ďalšie zvyšovanie počtu zvukov sa už neprejaví ako selekčná výhoda živočícha, môže dokonca viesť k zníženiu jeho fitnes (jeho

komunikácia sa v dôsledku zvýšeného výskytu chýb stáva zmätenou, pozri obr. 11). Prenos asociačnej matice $\boldsymbol{A}$ z rodičov na potomkov sa môže chápať ako kultúrny proces, t. j. prenos negenetickej informácie prostredníctvom sociálnej interakcie – učenia. Bolo ukázané, že presnosť tohto procesu je veľmi dôležitá pre evolučnú emergenciu koherentnej komunikácie medzi živočíchmi prostredníctvom asociácií zvuk – objekt. Ak výskyt chýb pri procese učenia presiahne určitú relatívne malú bariéru, potom evolučná emergencia komunikácie zanikne, vzniká len „zmätený“ komunikačný systém s množstvom synoným a homoným, ktorý je v podstate nezrozumiteľný.

5 Záver

Matematický pohľad na evolúciu jazyka je umožnený moderným prístupom založeným na kombinácii multiagentových systémov, teórii replikátorov a predstáv univerzálneho darwinizmu. V prvej etape sme študovali vlastnosti jednoduchého komunikačného systému, v ktorom sú objekty interpretované jednoduchými zvukmi. Bolo ukázané, že tento prístup obsahuje vnútorné limity, ktoré obmedzujú evolučnú výhodnosť prístupu pri snahe zvyšovať počet signálov. Tieto limity môžu byť čiastočne posunuté k vyšším počtom tak, že sa predpokladá vznik komplexných signálov spájaním elementárnych foném alebo vznikom „protogramatiky“, ktorá umožňuje kombinácie signálov popisujúcich objekty, resp. ich akcie. Všetky tieto prístupy poskytujú len konečné komunikačné systémy, ktorých ďalšie rozširovanie má evolučné limity, ďalšie zvyšovanie počtu signálov už nie je evolučne výhodné.

Až existencia gramatiky zabezpečuje ľudskému jazyku jeho komunikačnú neohraničenosť, možnosť vyjadrovať pomocou konečného počtu slov (zostrojených taktiež z konečného počtu foném) neohraničený počet rôznych situácií. Jej vznik sa dá modelovať podobným prístupom ako bol uvedený v tomto príspevku. Tento prístup si ale k úspešnej emergencii gramatiky vyžaduje doplnenie modelu o prvky kultúrneho prenosu v podobe mémov a Baldwinovho efektu [53]. Pre pokročilejšie modely vývoja jazyka sa tiež používa evolúcia neurónových sietí [35,50].

Poďakovanie: Tento príspevok vznikol za podpory grantovej agentúry VEGA SR v rámci grantovej úlohy 1/4053/07 a za podpory grantovej agentúry APVT v rámci grantovej úlohy APVT-20-002504.

Literatúra

[1] Baldwin, J. M.: A New Factor in Evolution. *American Naturalist* **30** (1896) 441-451.
[2] Belew, R. K., Mitchell, M. (eds.)*: Adaptive Individuals in Evolving Populations: Models and Algorithms.* Reading, MA: Addison-Wesley 1996.
[3] Bickerton, D.: Foraging Versus Social Intelligence in the Evolution of Protolanguage. In Wray, A. (ed.): *The Transition to Language*. Oxford: Oxford University Press 2002, chapter 10.
[4] Bickerton, D.: *Language and Species*. University Chicago, IL: Chicago Press 1990.
[5] Boyd, R., Richerson, P. J.: *Culture and the Evolutionary Process.* Chicago, IL: University of Chicago Press 1985.

[6] Briscoe, E. J.: Grammatical Assimilation. In: Christiansen, M. H., Kirby, S. (eds.): *Language Evolution: The States of the Art*. New York: Oxford University Press 2003, pp. 295-316.
[7] Cangleosi, A., Parisi, D.: *Simulating the Evolution of Language*. Berlin: Springer Verlag 2002.
[8] Cavalli-Sforza, L. L., Feldman, M. W.: *Cultural Transmission and Evolution: A quantitative approach.* Princeton, NJ: Princeton University Press 1981.
[9] Chomsky, N.: *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press 1965.
[10] Chomsky, N.: *Language and Mind*. New York: Harcourt Brace Jovanovich 1972.
[11] Dawkins, R.: *The Blind Watchmaker*. Oxford: Oxford University Press 1986.
[12] Dawkins, R.: *The Selfish Gen*e. Oxford: Oxford University Press 1976.
[13] Dennett, D.: *Darwin's dangerous ide*a. Hammondsworth: The Penguin Press 1995.
[14] Eigen, M., Schuster P.: The Hypercycles: A Principle of Natural Evolution. *Naturwissenschaften* **64** (1977) 541-565; **65** (1978) 7-41; **65** (1978) 341-369.
[15] Eigen, M.: Self organization of matter and the evolution of biological macro molecules. *Naturwissenschaften* **58** (1971) 465-523.
[16] Gilbert, N., Conte, R. (eds.): *Artificial Societies. The computer simulation of social life.* London: UCL Press 1995.
[17] Hinton, G. E., Nowlan, S. J.: How learning can guide evolution. *Complex Systems* **1** (1987) 495-502.
[18] Hofbauer, J., Sigmund, K.: *Evolutionary Games and Population Dynamics*. Cambridge, UK: Cambridge University Press 1998.
[19] Holland, J. H.: *Adaptation in Natural and Artificial Systems*. Ann Arbor, MI: University of Michigan Press 1975.
[20] Hurford, J. R.: Biological evolution of the Saussurean sign as a component of the language acquisition device. *Lingua* **77** (1989) 187-222.
[21] Jackendoff, R.: *Foundation of Language*. Oxford: Oxford University Press 2001.
[22] Jackendoff, R.: *The Architecture of the Language Faculty*. Cambridge, MA: MIT Press 1997.
[23] Jacob, F.: *Le Jeu des possibless, Essai sur la diversité du vivant*. Paris: Libraire Arthéme Fayard 1987.
[24] Komarova, N. L., Niyogi, P., Nowak, M. A.: The Evolutionary Dynamics of Grammar Acquisition. *J. Theoret. Biol*. **209** (2001) 43-59.
[25] Komarova, N. L., Nowak, M. A.: Language Dynamics in Finite Populations. *J. Theor. Biol.* **221** (2003) 445-457.
[26] Kubík, A.: *Inteligentní agenty.* Brno: Computer Press 2004.
[27] Kvasnička, V., Gál, E., Pospíchal, J.: Memetika "*in silico*". V Nosek, J. (editor), *Memy ve vědě a filozofii*, Praha: Filosofia, Nakl. FU AV ČR 2004, str. 101-119.
[28] Kvasnička, V., Pospíchal, J., Tiňo, J.: *Evolučné algoritmy*. Bratislava: Vydavateľstvo STU 2000.
[29] Kvasnička, V., Pospíchal, J.: An Emergence of Coordinated Communication in Populations of Agents. *Artificial Life* **5** (2000) 319-342.
[30] Kvasnička, V., Pospíchal, J.: Artificial Chemistry and Molecular Darwinian Evolution in Silico. *Collection of Czechoslovak Chemical Communications* **68** (2003) 139-177.
[31] Kvasnička, V., Pospíchal, J.: Replicator Theory of Coevolution of Genes and Memes. *Neural Newtorks World* **13** (2003) 255-266.

[32] Kvasnička, V., Pospíchal, J.: Simulation of Evolution of Dawkins Memes. *Evolution and Cognition* **5** (1999) 75-86.
[33] Kvasnička, V., Pospíchal. J.: Replicator Theory of Coevolution of Chomsky´s Universal Grammars and Generative Grammars. In: Peter Van Emde Boas, Pokorný, J., Bieliková, M., Štuller J. /eds./: *SOFSEM 2004: Theory and Practice of Computer Science. Proceedings Volume II*. Prague, Matfyzpress 2004, pp. 115-123.
[34] Kvasnička, V., Pospíchal. J.: Simulation of Baldwin effect and Dawkins memes by genetic algorithm. In Roy, R., Furuhashi, T. and Chawdhry, P. K. (eds.): *Advances in Soft Computing*. London: Springer-Verlag 1999, pp. 481-496.
[35] Kvasnička, V.: Konekcionistické modelovanie v kognitívnych vedách. *Filozofia* **58**(1) (2003) 35-43; Kvasnička, V., Pospíchal, J.: Evolúcia jazyka a univerzálny darwinizmus. In: Kognice a umělý život III. FPF SU, Ostrava, 2003, pp. 121-150.
[36] Maynard Smith J., Price, G.: The Logic of Animal Conflict. *Nature* **246** (1973) 15-18.
[37] Maynard Smith J.: *Evolution and the Theory of Games*. Cambridge, UK: Cambridge University Press 1982.
[38] Maynard Smith, J., Szathmary, E.: *The Major Transitions in Evolution*. Oxford: Freeman Spectrum 1995.
[39] Mitchener, G., Nowak, M. A.: Competitive Exclusion and Coexistence of Universal Grammars. *B. Math. Biol*. **265** (2003) 67-93.
[40] Niyogi, P.: The Computational Nature of Language Learning and Evolution. Cambridge, MA: MIT Press 2006.
[41] Nowak, M. A., Komarova, N. L., Niyogi, P.: Computational and evolutionary aspects of language. *Nature* **417** (2002) 611-617.
[42] Nowak, M. A., Komarova, N. L., Niyogi, P.: Evolution of Universal Grammar. *Science* **291** (2001) 114-118.
[43] Nowak, M. A., Komarova, N. L.: Towards an evolutionary theory of language. *Trends in Cog. Sci*. 5 (2001) 288-295.
[44] Nowak, M. A., Krakauer, D. C., Dress, A.: An error limit for the evolution of language. *Proc. Roy. Soc. London* **266** (1999) 2131-2136.
[45] Nowak, M. A., Krakauer, D. C.: The evolution of language. *Proc. Natl. Acad. Sci. USA* **96** (1999) 8028-8033.
[46] Nowak, M. A., Plotkin, J. B., Jansen, V. A. A.: The evolution of syntactic communication. *Nature* **404** (2000) 495-498.
[47] Nowak, M. A., Plotkin, J. B., Krakauer, D. C.: The Evolutionary Language Game. *J. Theoret. Biol*. **200** (1999) 147-162.
[48] Nowak, M. A.: From Quasispecies to Universal Grammar. *Z. Phys. Chemie* **216** (2002) 5-20.
[49] Nowak, M. A.: The Basic Reproductive Ratio of a Word, the Maximum Size of a Lexicon. *J. Theoret. Biol.* **203** (2000) 179-189.
[50] Oudeyer, P. Y.: *Self-Organization in the Evolution of Speech*. Oxford: Oxford University Press 2006.
[51] Pinker, S., Bloom, A.: Natural language and natural selection. *Behavioral Brain Sci*. **13** (1990) 707-784.
[52] Pinker, S.: Survival of the clearest. *Nature* **404** (2000) 441-442.
[53] Rybár, J., Beňušková, Ľ., Kvasnička, V. (editori): *Kognitívne vedy*. Bratislava: Kalligram 2002.

Spojení samoorganizace s výpočty: minimální život v moři umělých molekul

Jiří Wiedermann[1]

Abstrakt. Bakteroid je formální abstraktní hybridní systém, který ve své činnosti kombinuje výpočetní a nevýpočetní mechanizmy. Ukážeme, že v prostředí umělých molekul, nadaných jistými samoorganizačními schopnostmi, některé bakteroidy vykazují znaky minimálního života: jsou autonomní, replikují se a mají schopnost darwinovské evoluce. Návrh bakteroidů je inspirován představami současné molekulární biologie o dnes již neexistujících (či zatím neobjevených) formách protoživota.

„Maybe there's something beyond computation in the sense that we don't understand and we can't describe what's going on inside living systems using computation only…Maybe we're missing something, but what could that something be?"
Rodney Brooks, 2002 [2]

1 Úvod

Koncem roku 2002 získal Dr. Craig Venter, známý jako „luštitel" lidského genomu, a Dr. Hamilton Smith, laureát Nobelovy ceny, lékař, lukrativní grant amerického Ministerstva energetiky na projekt syntézy umělého života, resp. přesněji „minimální bakterie", která by byla schopna generovat „čistou" energii (produkcí vodíku), případně mírnit globální oteplování („požíráním" CO_2) [8]. Tento program stále pokračuje. V současné době byl „přečten" genetický kód stovek různých mikrobů s cílem nalézt principy, jak fungují nejjednodušší formy života a poté sestrojit příslušné počítačové modely. U té příležitosti se rozhořely debaty ohledně definice „minimálního života". I když se tyto debaty pohybovaly a pohybují především v rámci reálného života, přirozenou půdou pro takové debaty je i oblast umělého života (ALife).

Jedním z hlavních cílů umělého života je totiž studium zákonitostí života, které jsou nezávislé na mediu, v němž se odehrává. Takovým mediem může být počítač. Hlavní hnací silou přesvědčení, že (také) v počítačích může povstat život, je schopnost počítačů simulovat reálný svět molekul a komplexních chemických a fyzikálních interakcí mezi nimi. Jedna z možností syntézy ALife v počítači je tedy věrná simulace všech chemických procesů a fyzikálních zákonitostí – jakéhosi věrného obrazu našeho vesmíru – kterým podléhají reálné molekuly při vzniku života. To se děje v rámci tzv. umělé chemie (viz. např. [5]). Zde však narážíme na naše nedostatečné znalosti odpovídajících procesů a také na nedostatečnou kapacitu počítačů pro účely takové simulace (viz např. [20]).

Další možností je kvalifikovaný odhad těchto podrobností, sestrojení příslušného počítačového modelu a následné počítačové ověření, jestli i za takových ne zcela potvrzených předpokladů vznikne (teď už jistě umělý) život. Na základě takových motivací vznikla a stále vzniká řada počítačových modelů, které více či méně zdařile

[1] Ústav informatiky AV ČR, v.v.i., Pod Vodárenskou věží 2, Praha 8, E-mail: jiri.wiedermann@cs.cas.cz

modelují ne—li umělý život, tak alespoň některé jeho aspekty. Takovým aspektem může být např. schopnost zpracování informací ve smyslu turingovské vypočítatelnosti a míra takové schopnosti. Zde je základním modelem „praotec“ všech počítačů, Turingův stroj, který byl Turingem navržen také podle biologického vzoru člověka-počtáře. Jiný model – neuronové sítě – byl inspirován přirozenými neurony a jejich (tušenou) možností učit se.

Další, relativně pozdější modely, jsou celulární automaty[2], ještě pozdější vycházejí z různých gramatických modelů (L-systémy, eko-gramatické systémy). Tyto jsou těsně následované modely, které ještě více zdůrazňují motivaci molekulární biologií (molekulární výpočty, Watson-Crickovy automaty, spojovací systémy apod.). V poslední době jsou populární i tzv. membránové výpočty (viz např. [11]), které jsou zřetelně motivovány buněčnou molekulární biologií a přepisovacími systémy. Nejnovějším hitem jsou tzv. samosestavovací procesy (self-assembly, viz. např. [1],[16],[19]), inspirované samo-organizačními schopnostmi biomolekul. Příznačné pro teoretické studium všech zmíněných modelů je odklon od původní biologické motivace v okamžiku, kdy se ukazuje jejich „praktický význam“. Jako typický příklad mohou posloužit P-systémy, kde se zdůrazňuje, že cílem těchto modelů není modelování celulárního života, ale naopak zkoumání výpočetní síly modelů inspirovanými celulární biologií [11], anebo v případě samosestavování i využití těchto procesů v (nano)technologii. Tím se ovšem dostávají mimo rámec umělého života, protože např. výpočetní univerzalita v Turingovském smyslu, která se zdá býti svatým grálem těchto a podobných modelů, není fundamentální vlastností živých systému. Naopak, schopnost samoreprodukce a darwinovské evoluce – základních podmínek života – se v rámci těchto modelů většinou nezkoumají, i když i zde se najdou výjimky (viz např. [17]).

Náš příspěvek si klade za cíl zkoumat zejména z výpočetního hlediska přesně to, čemu se shora zmíněné modely, byť inspirované živými systémy, většinou vyhýbají: vlastnosti a projevy života, hledání resp. modelování minimálních vlastností nutných pro život. Takovému životu se říká minimální život. Je třeba zdůraznit již teď, že naším cílem také není modelování reálného života či našich představ o něm – necháme se jimi pouze inspirovat. V důsledku toho zůstaneme pevně na půdě ALife. V současné době mezi odborníky nepanuje shoda o definici života, zachycující jeho rozmanitost v plné šíři. S minimálním životem je to jiné, zde převládající a pro nás použitelná definice minimálního života, je tato [13]: Nějaký systém lze považovat za živý, když

- je autonomní (tj. nezávislý na přítomnosti produktů jiného života v prostředí)
- reprodukuje se
- a vyvíjí se v souladu s darwinovskou evolucí

Výchozím výpočetním modelem, pomocí kterého budeme zkoumat minimální život, pro nás bude tzv. asynchronní multitransducer s přepínacími stavy. Bude to multimnožina konečných automatů (přesněji řečeno: automatů s výstupem, tj. transducerů), které mohou navzájem asynchronně komunikovat prostřednictvím stavů, které se navíc mohou přepínat z počátečních na ne-počáteční a naopak. Tato množina automatů zpracovává multimnožinu symbolů a řetězců nad nimi, které se nacházejí v prostředí,

[2] Vzhledem k relativní známosti těchto i dalších v úvodě zmiňovaných modelů v komunitě odborníků umělé inteligence a umělého života, s výjimkou nejnovějších modelů neuvádíme žádné reference na příslušnou literaturu.

způsobem podobným, jako pracují P-systémy, tj. každý automat jako by aktivně vyhledal ten řetězec, pro zpracování kterého je naprogramován (pokud takový řetězec v okolí existuje). Automaty multitransduceru „drží pohromadě“ pomocí membrány, ve které se všechny nacházejí a která je produktem činnosti některých automatů. Membránou jsou tedy jednotlivé multitransducery, které se mohou lišit ve své činnosti a v konečném důsledku i „morfologii“, od sebe fyzicky oddělené. Multitransducer umístněný v membráně se nazývá bakteroid.

Dále multitransducer zabezpečí kopírování (reprodukci) „genetické informace“ bakteroidu, která se nachází uvnitř membrány také ve formě řetězce symbolů a popisuje vlastně přechodové (řídící) funkce jednotlivých automatů. Vlastnost genetické mutace je dosažena povolením (resp. vyžadováním) přepisovacích chyb při kopírování genetické informace. Jiné automaty mají na starosti přípravu (syntezi, modelovanou jako výpočet) stavebních kamenů různých dalších percepčních a motorických orgánů bakteroidů. Tyto orgány však multitransducer nekonstruuje – pouze připravuje jejich součásti. To znamená, že multitransducer nemá k dispozici „návod“ na konstrukci těchto orgánů, pouze „v pravý čas“ dodá součástky potřebné pro další fázi vývoje orgánu. Vlastní asembláž je považována za nevýpočetní proces (nezaměňovat s nevypočitatelným procesem, pro který neexistuje algoritmus) v tom smyslu, že není řízena multitransducerem.

Kýžené orgány (receptor, motor, transpoziční kanál – jakási „ústa“ bakteroidu, atp.) pak vznikají samoorganizací, resp. „samosestavováním“, tj. pomocí vlastnosti, kterou a priori postulujeme pro některé typy objektů, které automaty zpracovávají. Výsledný mechanizmus bakteroidů tudíž kombinuje ryze výpočetní mechanizmus multitransduceru s „nevýpočetním“ mechanismem pro konstrukci membrány a percepčních a lokomočních orgánů, který je založen na vlastnostech samoorganizace příslušných součástí. Tento nevýpočetní mechanizmus do jisté míry připomíná orákula, která se ve světe Turingových strojů používají za účelem získání výsledků, které daný mechanismus není principielně schopen obdržet vlastními silami. Jiný pohled na věc je, že výpočetní mechanizmy zabezpečují epistemickou složku práce bakteroidu, kdežto nevýpočetní mechanizmy složku ontickou [6]. Poslední vlastnost bakteroidů, totiž množení, je zabezpečeno pomocí dělení membrány na dvě části, v jedné z nich se nachází původní a ve druhé zkopírovaná a případně zmutovaná genetická informace. Samotný mechanizmus dělení membrány je opět nevýpočetní a pracuje na základě využití fyzikálních a chemických zákonů, panujících v prostředí, ve kterém se bakteroid nachází.

Z hlediska ALife je přínosem práce již samotný návrh bakteroidu, který striktně odděluje výpočetní aspekty jeho činnosti od nevýpočetních. Získané výsledky ukazují důležitost komunikace mezi jednotlivými automaty multitransduceru a dávají představu o typu a rozsahu epistemické práce bakteroidu. Komunikace se v našem případě modeluje pomocí přepínání stavů a zhruba odpovídá aktivaci genů v genetické informaci organizmu. V práci dále ukážeme, že definice bakteroidu je „dobrá“ tím způsobem, že zkonstruujeme jednoduchý bakteroid – tzv. protobakteroid – o kterém dokážeme, že vykazuje známky minimálního života. Již tento výsledek samotný umožňuje hlubší vhled do principů minimálního života a úlohy zpracování informací ve vývoji protoorganismů na straně jedné a úlohy samoorganizace na straně druhé. Další evolučně vyšší typ bakteroidu, tzv. selektivní bakteroid, již vykazuje známky kognice – dovede rozlišovat mezi dvěma typy „potravy“ a upřednostnit výživnější stravu. Konečně nejvyvinutější typ bakteroidu, jehož konstrukci zde nastíníme, je mobilní bakteroid vykazující známky

inteligentního chování – pohybuje se v směru gradientu stravy. Dva posledně zmíněné bakteroidy jsou příspěvkem k pochopení vzniku a fungování kognice a inteligentního chování.

Článek pozůstává ze sedmi částí. První část obsahuje úvod do problematiky a nástin hlavních výsledků a jejich přínosů. Dále následuje definice multitransduceru, ve třetí části popis jeho „vtělení“ do membrány a popis mechanizmů růstu a dělení membrány. V části 4 definujeme pojem bakteroidu a jeho speciálního případu protobakteroidu, který lze považovat za nejjednodušší bakteroid splňující podmínky minimálního života. Toto dokážeme v 5. části. Dále v 6. části popíšeme „evoluční inovace“ protobakteroidu. Základem těchto inovací je tzv. selektivní bakteroid, který je schopen preferovat kvalitnější stravu v prostředí bohatém na živiny. Tento bakteroid tudíž vykazuje elementární kognitivní schopnosti a za tím účelem musí být vybaven perceptory a mít schopnost činit rozhodnutí ve prospěch lepší alternativy. Další typ bakteroidu je již vybaven i pohybovým aparátem a dovede se pohybovat směrem ke větší koncentraci potravy. V závěrečné sedmé části jsou zhodnoceny přínosy práce a nastíněn další postup výzkumu.

Článek má předběžný referující charakter popisující problematiku ve stavu zrodu. Zatím není ustálen ani model, ani formalizmus, ani terminologie, výsledky se pouze rýsují.

2 Interaktivní konečný multitransducer

Soustřeďme se nejprve na výpočetní aspekt zamýšleného formálního systému, který má modelovat mechanizmy zpracování informací a jejich využití pro řízení činnosti umělého organizmu. Pro tyto účely se jako první nabízí využití konečných automatů. Jejich modifikace pro účely interaktivního zpracování (tj. zpracování potenciálně nekonečného proudu předem neznámých vstupních dat) je poměrně přímočará záležitost a výsledné zařízení – tzv. interaktivní konečný transducer – bylo popsáno již mnohokrát (viz např. [14]). Pro naše záměry budeme potřebovat ještě jiné, byť podobné, zařízení – tzv. multitransducer.

Na multitransducer lze pohlížet jako na multimnožinu Mealyho automatů konečně mnoha typů, které pracují asynchronně. To znamená, že i když je popis multitransduceru (a tudíž všech automatů dohromady) konečný, mohutnost jejich multimnožiny je prakticky neomezená a počet automatů v ní závisí od množství objektů, které automaty mají v daném chvíli k dispozici pro zpracování (viz v dalším popis činnosti multitransduceru). Z hlediska výpočtů se tedy jedná o vysoce paralelní zařízení. Každý typ Mealyho automatu je popsán svou vlastní přechodovou funkcí a je upraven pro případ zpracování vstupů, které nejsou předem dány na pásce. Předpokládáme, že multitransducer se nachází v prostředí tvořeném multimnožinou řetězců vstupních a výstupních symbolů, kterým budeme v našem kontextu říkat objekty. Také tato multimnožina není předem daná, může se časem měnit, tj. mohou do ní přibývat nové typy objektů anebo se může měnit multiplicita výskytu jednotlivých objektů.

Multitransducer pracuje tak, že soustavně transformuje předepsaným způsobem objekty, nacházející se v prostředí, na jiné objekty. Vstupní objekty jsou čteny přes tzv. vstupní porty a výstupní objekty jsou vypisovány přes výstupní porty jednotlivých automatů. Pokud je objekt strukturovaný, pozůstává z více symbolů, tak je čten

sekvenčně, jako řetězec. Každý automat multitransduceru může být navržen i naprogramován i tak, že zpracovává více objektů současně, podobně jako klasický vícehlavý konečný automat. Přitom předpokládáme, že každý automat čte objekty z prostředí selektivně – má zvláštní schopnost „najít“ a zpracovat v prostředí ten objekt, pro který je naprogramován, pokud v prostředí takový objekt existuje. Tato vlastnost automatů může být chápána i jako vlastnost prostředí, ve kterém se multitransducer nachází – prostředí se jakoby pokouší zpracovat každý objekt každým z automatů, ze kterých se multitransducer skládá, a pokud taková dvojice automat – objekt existuje, tak se zpracování objektu automatem uskuteční.

Prostředí má tedy potenciál pro realizaci vysoce paralelních výpočtů. V případě, kdy je možné daný objekt zpracovat dvěma různými automaty, tak se náhodně vybere jeden z nich. Pokud v prostředí takový objekt, který by mohl automat zpracovat, neexistuje, automat čeká ve stavu, ve kterém předtím skončil. Jakmile se v prostředí znovu objeví objekt vhodný ke zpracování, automat se opět spustí z počátečního stavu. Objekt „přečtený“ na vstupním portu po této akci „zmizí“ – je transformován na nějaký výstupní objekt. Poté je port připraven číst další objekt.

Multitransducer se od „běžné“ množiny automatů liší v jednom podstatném bodě – množina počátečních stavů jednotlivých automatů není pevná, předem daná, neměnná, ale jednotlivé stavy se v závislosti na průběhu výpočtu mohou stát počátečními stavy a existující počáteční stavy mohou vlastnost „počátečnosti“ ztratit. Příkazy pro aktivaci počátečních stavů jsou součástí přechodové funkce automatů. Počáteční aktivace stavů je popsána ve specifikaci multitransduceru. Stavy, které jsou momentálně počáteční, budeme nazývat i aktivními stavy. Dále popíšeme formálně případ multitransduceru s jednonásobnými vstupními porty, tj. multitransduceru, ve kterém každý automat čte své vstupy právě prostřednictvím jednoho portu.

DEFINICE: *Interaktivní asynchronní konečný multitransducer* $T = (I, O, S, B, \delta)$ s jedno-násobnými vstupními porty je pětice, kde

- I a O jsou konečné abecedy objektů, I je abeceda vstupní a O je abeceda výstupní
- S je konečná abeceda stavů
- $B \subseteq S$ je množina aktivních stavů na počátku výpočtu
- δ je přechodová funkce tvaru $I \times S \rightarrow S \times O \times S \times \{0,1\}$, která pro objekt $v \in I$ čtený na vstupním portu zařízení a stav $s \in S$ přiřadí nový stav $r \in S$, na výstupní port pošle objekt $w \in O$ a stavu q přiřadí hodnotu aktivace 0 nebo 1 (0 znamená, že stav není počáteční, 1 opak); tomu odpovídá zápis $\delta(v,s) = (r,w,q,0)$, resp. $\delta(v,s) = (r,w,q,1)$.

Dále se zdá býti rozumným požadovat, aby multitransducer nemohl transformovat prázdný řetězec na neprázdný řetězec a naopak, tj. multitransducer nemůže „tvořit“ něco z ničeho, ani udělat z něčeho nic.

U tohoto zařízení se nepočítá s jeho evolucí – změnu jeho činnosti lze dosáhnout pouze změnou jeho přechodové funkce. Je zřejmé, že aktivita jednotlivých stavů je na začátku výpočtu daná množinou B a dále se mění v závislosti na průběhu výpočtu, tj. v závislosti na zpracovávaných objektech a jejich pořadí (nezapomínejme, že jednotlivé automaty pracují asynchronně).

3 Vtělení interaktivního transduceru

V práci [15] později i v [18] bylo navrženo rozšíření konečných kognitivních transducerů na tzv. konečné kognitivní agenty, kteří se liší od dříve zmíněných zařízení tím, že součástí modelu jsou i tzv. percepčně-motorické jednotky, které umožňují interakci agenta s okolím a představují tak jakési první přiblížení agentova těla. Tento model však nemá žádné mechanizmy pro modelování „růstu“ těla, ani agentovy evoluce, jak „tělesné“ anebo „duševní“, které se zdají nevyhnutelné, pokud chceme modelovat vznik a vývoj jednoduchých forem života. V tomto pojetí mentální evoluce znamená samozřejmě změnu mechanizmů, řídících zpracování dat agentem – tj. změnu přechodové funkce odpovídajícího transduceru. To jsme částečně dosáhli v předchozím modelu multitransduceru pomocí přídavného mechanizmu aktivace stavů. V dalším k tomu přibude ještě mutace – náhodná změna přechodové funkce multitransduceru. Zbývá vyřešit problematiku „vtělení“ (ve smyslu teorie tzv. vtělené umělé inteligence – viz. např. [12]) programovatelného kognitivního transduceru (či agenta) do vhodné tělesné schránky, abychom mohli zachytit i aspekty tělesného vývoje. Tato schránka tudíž musí plnit několik základních úkolů: musí chránit řídící a informační mechanizmy agenta před vlivy vnějšího světa, musí být „nosičem“ percepčně-motorických jednotek agenta, a v neposlední řadě musí umožnit vývoj (zejména růst) a dělení (množení) agenta.

Jako model nejjednoduššího „těla“ budeme uvažovat (třírozměrnou) *membránu*. Budeme ji modelovat jako třírozměrnou strukturu, pozůstávající ze speciálních objektů, tzv. *destiček*. Tyto destičky mají za jistých fyzikálně-chemických okolností vlastnost „samoorganizace“ do kulovité struktury, kterou budeme nazývat membrána. To znamená, že za příznivých okolností jakýkoliv shluk takových destiček, které jsou dostatečně blízké jedna druhé, se spontánně uspořádá do kulovité struktury a navíc, jestli se v blízkosti takové membrány nachází jedna anebo více dalších destiček, tak se tyto inkorporují do existující membrány. Za jiných okolností, např. pokud se membrána „naplní k prasknutí“, se membrána „roztrhne“ – rozdělí na dvě přibližně stejné části, které se opět spontánně uzavřou do kulovitého tvaru. Původní obsah membrány se přitom také rozdělí přibližně napůl, malé ztráty obsahu jsou však možné a pravděpodobné. Vlastnost růstu a dělení membrány lze vidět jako důsledek fyzikálních a chemických zákonů, působících v daném prostředí na materiál membrány. Rychlost růstu membrány bude (kromě jiného) záviset na přísunu materiálu (destiček). V dalším uvidíme, že destičky mohou vznikat i za přispění automatů, jako výsledek zpracování jiných objektů, které se nacházejí v prostředí a nemají vlastnost samoorganizace. V membráně se dále mohou nacházet percepční a motorické jednotky agenta. Materiál na jejich konstrukci dodávají automaty umístněné uvnitř membrány. Přítomnost těchto jednotek a jejich počet závisí od přísunu materiálu, ze kterých jsou budovány.

Všechny automaty multitransduceru se řídí programem, který se po vzoru DNA (resp. prokaryotických bakterií) nachází uvnitř membrány a má i podobný tvar ve formě (svinuté) konečné pásky. Z hlediska multitransduceru je to objekt jako kterýkoliv jiný a proto se může stát předmětem zpracování některého z automatů, které tvoří multitransducer. Je to vlastně lineární zápis specifikace multitransduceru. V prvním přiblížení si tento zápis lze představit jako posloupnost instrukcí „co dělat, když na vstupu nějakého automatu ve stavu s je objekt v“. Tento objekt je potom zpracován způsobem odpovídajícím přechodové funkci. Jednotlivé segmenty zápisu přechodové funkce

odpovídají zápisu jedné instrukce v přechodové funkci, tj. zobrazení tvaru $I \times S \rightarrow S \times O \times S \times \{0,1\}$ pro konkrétní hodnoty parametrů. Na program lze nahlížet jako na „návod“ (instrukce) pro „výrobu“ různých objektů, které lze dále použít jako jakési prefabrikáty např. při stavbě membrány anebo percepčních a motorických orgánů. Dále program obsahuje instrukce pro započetí jejich konstrukce, případně pro jejich zpětné rozložení.

To se děje pomocí aktivace stavů do té doby nečinných automatů anebo naopak pomocí deaktivace stavů činných automatů. Pozornosti hodným a důležitým je případ, když se aktivuje automat, jehož úkolem je vytvořit kopii objektu, který reprezentuje program multitransduceru. Takový automat má dva vstupy – jedním čte stávající program a druhým přijímá „materiál“ (objekty), ze kterého na výstupu konstruuje kopii programu. Tj. jedná se o jednoduchý kopírovací automat. Zdá se, že někteří autoři se domnívají, že pro modelování sebereprodukce živých organizmů je nutné, aby součástí genetického popisu organizmu byl kromě popisu reprodukčního zařízení pro kopírování genetické informace také „návod“, jak se „sebereprodukovat“ (to je např. esence známého von Neumannova výsledku o sebereprodukujících se automatech). V naší analogii by to znamenalo, že součástí multitransduceru by měl být i popis, jak sestrojit membránu a kdy a jak ji rozdělit, jak zabezpečit, aby v každé kopii organismu byla právě jedna kopie genetické informace, atd. To je sice myslitelné, ale pro všechny další účely pravděpodobně zbytečné a z hlediska evoluce dost složité a tudíž málo pravděpodobné. Mechanismus kopírovaní DNA v reálných buňkách ukazuje, že pro kopírování genetické informace stačí právě popsaný kopírovací automat.

Při své činnosti multitransducer „spoléhá“ na samoorganizační schopnost některých objektů, jako jsme již viděli v případě destiček. Tato vlastnost objektů je inspirována vlastnostmi některých organických molekul, které se za jistých specifických fyzikálně-chemických podmínek spontánně organizují do větších, složitějších struktur, přičemž volná energie systému klesá. Molekuly přitom není potřebné nijak zvláště instruovat, jakým způsobme to mají činit. V našem případě budeme takové vlastnosti vybraných objektů pouze postulovat, podobně jako podmínky, za kterých tak činí. Nebudeme se nutně starat o to, zda takové objekty ve skutečnosti existují a jak fungují. Pokud tedy multitransducer vygeneruje z jiných objektů objekty, ze kterých je možné „poskládat“ pomocí samoorganizace nějaký percepční či motorický orgán, tak takový orgán se z těchto nově vygenerovaných objektů vytvoří „sám od sebe“ a bude umístněn v prostředí, které je *de facto* určeno vlastnostmi sebesestavovacích objektů, ze kterých se orgán skládá (např. stane se součástí membrány, jako v případě percepčních či lokomočních orgánů). Řízení procesu generace jednotlivých objektů, tj. rozhodnutí, kdy (za jakých okolností) co (jaké orgány či jejich součástky) generovat a z jakých objektů má na starosti multitransducer, popsaný výše zmíněným programem, který tvoří zvláštní objekt uvnitř membrány.

Pravděpodobně bude na místě pár slov o reálnosti navrhovaných výpočetních a nevýpočetních mechanismů vtělených multitransducerů. Výpočetní mechanismy jsou vlastně podstatným zjednodušením tzv. centrální dogmy molekulární biologie [3], totiž že každý gen v DNA molekule nese informaci potřebnou pro konstrukci jednoho proteinu, který ve formě enzymu řídí jednu chemickou reakci v rámci buňky. V našem modelu genu odpovídá jeden příkaz přechodové funkce multitransduceru a příslušný automat odpovídá zmíněnému proteinu. Chemickým sloučeninám (biomolekulám) odpovídají objekty,

reakcím odpovídají výpočty nad objekty. Objekty jsou tedy skutečnou fyzickou reprezentací dat, jsou jejich „ztělesněním“; konstrukcím složitějších objektů odpovídají výpočty anebo samoorganizace.

Dále, v reálném životě je DNA molekula „navinutá“ na skupinu proteinů, tzv. histonů. Přepínání genů se ve skutečných buňkách pravděpodobně děje pomocí chemických přepínačů (viz tzv. operonovou hypotézu), které „zakryjí“ část DNA s požadovaným genem a tím jej vyřadí z činnosti. Komunikace v rámci buňky funguje pomocí chemických či jiných signálů (změnou fyzikálních vlastností molekul indukovanými změnou jejich tvaru). Samoorganizace vybraných typů biomolekul (např. amfifilních molekul do tvaru třírozměrných obálek – viz např. [3]) je skutečností, která je experimentálně ověřena. Prostupování vstupních objektů přes membránu, případně přes specializované kanály či póry, nacházející se v membráně, také odpovídá skutečnosti, v praxi je to důsledek brownova pohybu molekul a chemicko-fyzikálních vlastností zúčastněných molekul (jak v membráně, tak i těch prostupujících). Více o samoorganizaci, její biochemické realizaci a jejím využití viz např. v přehledovém článku [19]. V teorii se zatím uvažují samosestavovací objekty čtvercového tvaru, ze kterých však nelze poskládat orgány, které potřebujeme; nicméně již tyto výsledky ukazují, že samosestavování objektů je v principu možné a že výsledný tvar složitějších objektů závisí na tvaru součástí, ze kterých se komplex skládá.

Nyní, když máme dostatečnou představu o mechanizmech jednotlivých součástí vtěleného multitransduceru, můžeme přikročit k popisu tzv. bakteroidu, který představuje jednoduchý umělý živý organizmus, žijící v oceánu objektů, ze kterých lze sestavit složitější objekty se samoorganizačními schopnostmi.

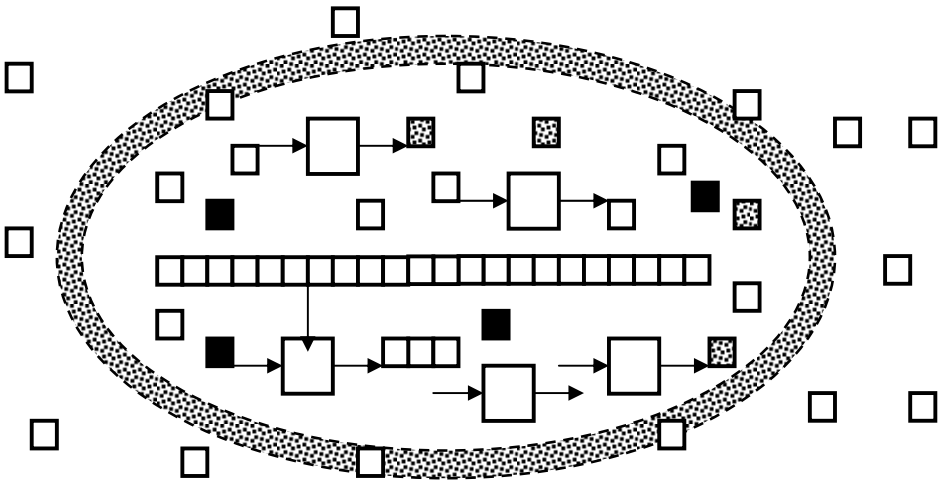
4 Bakteroid

Bakteroid je generický název pro dvojici, skládající se z multitransduceru a membrány, která multitransducer obklopuje. Podobně jako v případě skutečných bakterií se jednotlivé druhy bakteroidů liší svojí „výbavou“ – a to jak genetickou (zhruba bychom mohli říci, složitostí multitransduceru, který bakteroid řídí, anebo, řečeno jazykem evoluční biologie, genotypem), tak i fyzickou (fenotypem, tj. jak bakteroid „vypadá“, jakou má membránu, jaké má vlastnosti tato membrána, jaké má orgány, kde umístněné apod.). Nejjednodušší bakteroid v tomto směru je tzv. *protobakteroid*, který nemá žádné percepční a motorické orgány, skládá se pouze z membrány, která uvnitř obsahuje objekt, ve kterém je zakódován řídící program multitransduceru (viz schéma bakteroidu na obr. 1). Tomuto objektu budeme říkat jednoduše páska. Důvod pro existenci takového seskupení je v tom, že multitransducer syntetizuje z objektů, které jsou vně protobakteroidu a volně prostupují membránou, destičky, které vznikají uvnitř membrány a samoorganizací se inkorporují do již existující membrány.

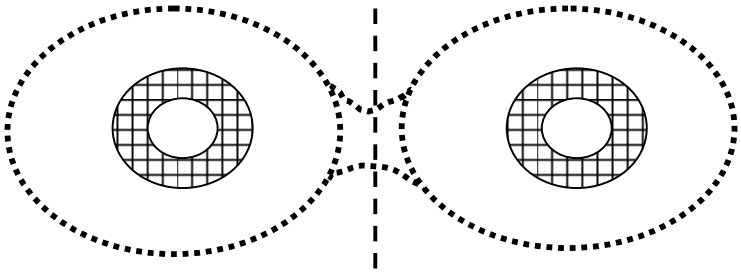
Příslušné automaty budeme nazývat *destičkové automaty*. Tím způsobem membrána roste. Současně s tím probíhá reprodukce programu multitransduceru, pomocí *kopírovacího automatu*, také z objektů, které volně prostupují membránou. Jak již bylo zmíněno dříve, předpokládejme, že původní i nová páska jsou svinuty do dvou kulovitých klubek. Dále přepokládejme, že membrána obepínající obě klubka není napnutá, ale že „vyrostla“ více než by stačilo k těsnému obepnutí klubek. To znamená, že výsledný tvar celého bakteroidu se čím dále tím více začne podobat jakési třírozměrné osmičce

a v okamžiku, kdy se vnitřní část membrány mezi oběma klubky pásky dostatečně přiblíží, zapracuje samoorganizační schopnost destiček a protobakteroid se rozdělí na dva více méně stejné protobakteroidy (viz. obr. 2)[3].

I z tohoto neformálního popisu je vidět, že protobakteroid potřebuje pro svou existenci dle výše zmíněných představ řadu objektů s různými vlastnostmi, z nichž musí být schopen generovat další objekty (např. destičky) ještě s jinými vlastnostmi, zejména samoorganizačními, které se opět organizují do dalších struktur (např. membrány), které mají ještě jiné vlastnosti, než její součásti a podléhají fyzikálním zákonům.



Obrázek 1. Schéma bakteroidu



Obrázek 2. Dělení bakteroidu

[3] V práci [6] se navrhuje jiný mechanismus dělení membrány inspirovaný reálnými bakteriemi: autoři přepokládají, že původní i zkopírovaná páska s genetickou informací se „připevní“ na protilehlé strany membrány a membrána se rozdělí tak, aby se při dělení obě pásky rozmotaly a oddělily a současně aby každá zůstala ve „své“ půlce membrány, která se samozřejmě současně uzavírá. Toto nedovedeme v našem modelu napodobit – problém je, jak určit „protilehlé“ strany pomocí mechanizmů konečných automatů (anebo jinak, pomocí nějakých fyzikálních vlastností) a zabezpečit dělení membrány právě mezi takto orientovanými póly.

Pro úplnost se pokusme nyní ještě blíže a formálněji specifikovat náš protobakteroid, částečně ve smyslu definice multitransduceru a s přihlédnutím na další vlastnosti jednotlivých objektů, nutné pro činnost protobakteroidu dle výše zmíněných představ.

Vstupní objekty: to jsou objekty, které se budou nacházet vně i uvnitř těla bakteroidu, ale nejsou součástí membrány; z nich se budou generovat *destičky* a z nich se bude sestavovat *kopie pásky* s řídícím programem transduceru.

x,y,z: objekty pro budování destiček a elementů pásky

Výstupní objekty: jsou výsledkem zpracování vstupních objektů automaty; z výstupních objektů typu *destička* se buduje samoorganizací objekt *membrána*; z dalších se buduje kopie pásky; membrána je prostupná pro všechny vstupní objekty směrem dovnitř a neprostupná pro výstupní objekty (směrem ven).

a,g,t,c: objekty, ze kterých se skládá páska;

d: destička

Multitransducer: skládá se ze tří automatů: *destičkového*, který syntetizuje destičky d z objektů x a y, *syntetizačního* automatu, který syntetizuje elementy a,g,t,c, z objektů z,v,w a *kopírovacího automatu*, který čte objekt p, skládající se z objektů a,g,t,c, a syntetizuje jeho kopii z objektů a,g,t,c.

p: páska s popisem přechodové funkce multitransduceru

Přechodová funkce: od jejího popisu upustíme, zatím není příliš zajímavá, protože nevyžaduje žádnou změnu aktivace počátečních stavů – destičkový automat i syntetizační automat jsou jednokrokové, stále se nacházejí v počátečním stavu. Cyklus kopírovacího automatu musí ovšem proběhnout pouze jednou; aby se páska nekopírovala podruhé, musí být dosažení okamžiku dělení synchronizováno s ukončením procesu kopírování. Za tím účelem nemá protobakterie žádné zvláštní mechanizmy, tato vlastnost musí být zaručena relativními rychlostmi všech zúčastněných automatů.

Další předpoklady: páska je svinutá do klubíčka, kopírování je ukončeno těsně předtím, než se membrána rozdělí výše popsaným mechanismem

Výše uvedené představy o fungování protobakteroidu jsou zajisté zcela jiné než představy ryze výpočetní a zde se rýsuje odpověď na Brooksovu otázku z citátu v úvodu této práce. Tyto představy však nepřekračují hranice našeho vědění o vlastnostech biomolekul a naopak jsou plně v souladu se současnými představami o prvních formách života a o fungování živých buněk (viz např. [4][13]) a téměř v dosahu možností současných technologií [1].

5 Protobakteroid a minimální život

Dále se budeme věnovat otázce, nakolik protobakteroid odpovídá současným představám o tzv. *minimálním životě*. Je to představa živého organizmu, který má minimální množinu vlastností takových, abychom ho mohli považovat za živý systém. Odebráním kterékoliv z těchto vlastností přestane organismus odpovídat našim představám o životě. Zdá se, že v současné době převažující operační definice vyžaduje splnění dvou podmínek [13]: příslušný systém musí mít v daném prostředí schopnost autonomní replikace a musí podléhat pravidlům darwinovské evoluce. Autonomní replikace znamená nepřetržitý růst a množení systému; obě tyto akce jsou závislé pouze na vstupních jednoduchých molekulách a nezávislé na produktech jiných forem života. Darwinovská evoluce

vyžaduje, aby se genetická variace (mutace) projevovala ve fenotypu organizmu a jeho chování a tím ovlivňovala přežívání a reprodukci organizmu.

Nuže, z návrhu protobakteroidu je vidět, že tento v daném prostředí, při dostatečném přísunu vstupních objektů, přežívá a množí se. Přitom jsme předpokládali, že vstupní objekty jsou jednodušší objekty, než výstupní, protože posledně zmíněné jsou „vypočítány“ z prve zmíněných. Darwinovskou evoluci zabezpečíme „malou úpravou“ přepisovacího (kopírovacího) mechanizmu: připustíme (resp. dokonce „zařídíme“), aby tento mechanismus byl nespolehlivý. To znamená, při kopírování mohou nastávat chyby nejenom „přepisovacího“ typu, ale třeba takové, že se některá část programu vůbec nezkopíruje, anebo se naopak zkopíruje dvakrát. Tím samozřejmě dochází ke „genetické variaci“ – program multitransduceru mutuje. Některé změny se vůbec neprojeví (třeba duplikace některých částí programu), jiné se projeví třeba okamžitě nefunkčností protobakteroidu anebo naopak jeho zlepšenou funkčností. Právě posledně zmíněné změny jsou ty změny, které znamenají „posunutí vývoje“. Je zajímavé si uvědomit, že podobně jako u živých organismů, i v našem modelu zásadnější pokrok nemůže nastat často za sebou, s malými časovými odstupy. Je to proto, že některé změny se musejí „naakumulovat“. Např. se některá část kódu duplikuje a zatím se v životě protobakteroidu neprojevuje. Dále může proběhnout mutace, která tuto část kódu vyřadí vůbec z činnosti (přepnutím počátečního stavu do pasivního stavu). Dále v této části mohou proběhnout další náhodné mutace a když se náhodou opět tato část kódu „zapne“, začne se protobakteroid chovat v určité situaci jinak, třeba efektivněji, než jeho předchůdci. To znamená, že kromě jakési pomalé, „plíživé evoluce“, postupující v malých, sotva pozorovatelných krůčcích, existuje i radikální evoluce propukající s delšími časovými prodlevami, ale projevující se zásadnějšími změnami jak ve fenotypu, tak i v chování protobakteroidu. Zdá se, že mechanizmus, který podporuje radikální inovaci způsobem popsaným výše, lze „zabudovat“ do kopírovacího automatu, takže vlastně může vzniknout „sám od sebe“ pomocí evoluce.

V dalším si popíšeme příklad bakteroidu, který je výsledkem takové radikálnější evoluce protobakteroidu uvažovaného v 4. části.

6 Selektivní bakteroid

Protobakteroid, byť vykazoval známky minimálního života, byl poměrně jednoduchý. Hlavní „oživovací trik“ spočíval v tom, že jeho „tělesný vývoj“, představovaný růstem membrány a kopírovacím mechanismem programu, byl řízen multitransducerem. Tím se dosáhlo, že „přežití“ protobakterie záviselo na souhře mezi programem, „kvalitou“ tohoto programu, přísunem materiálu na výrobu destiček a částí kódu a jejich výrobou. Ostatní věci zabezpečila samoorganizace a fyzikální zákony. Pokud potřebná souhra nebyla dosažena, protobakterie jednoduše nefungovala a nemnožila se. V opačném případě se množila a pokud mutací vzniklo nějaké vylepšení, které resultovalo např. v rychlejší růst protobakteroidu, tak takto modernizovaný bakteroid po čase v populaci protobakteroidů převládnul. Pro svou existenci protobakteroid nepotřeboval informace o vnějším prostředí. V dalším uvedeme příklad tzv. selektivního bakteroidu, který už bude využívat informace o prostředí a tím dosáhne v porovnání s protobakteroidem evoluční výhodu rychlejšího růstu. Tento příklad současně poslouží i jako ilustrace účinnosti mechanizmu průběžné

aktivace a deaktivace jednotlivých stavů, čím se zvýrazní i přínos tohoto mechanizmu pro studium genetické expresivity bakteroidů.

Uvažujme bakteroid, který se nachází v prostředí nutrientů, které jsou modelovány objekty typu A a B. Bakteroid může zužitkovat oba typy potravy, avšak potrava typu A je mnohem výživnější než potrava typu B – např. bakteroid může vyprodukovat z potravy typu A několikrát více deštiček a částí kódu než z potravy typu B. Protobakteroid, popsaný ve 4. části – říkejme mu $\mathcal{P}$ – může fungovat v takovém prostředí beze změny – nemusí nijak rozlišovat mezi jednotlivými nutrienty. Avšak bakteroid, který by v prostředí preferoval živiny typu A (pokud by těchto byl dostatek), by zajisté byl evolučně úspěšnější než $\mathcal{P}$. Bakteroid takového typu budeme označovat jako selektivní bakteroid.

Selektivní bakteroid – označme jej $\mathcal{S}$ – tedy musí nějakým způsobem zjistit, jestli je v současné době v okolí dostatek anebo nedostatek živin typu A a zařídit se podle toho. Jinými slovy, jestli je koncentrace živin typu A dostatečná, měl by $\mathcal{S}$ přestat přijímat živiny typu B, v opačném případě by měl akceptovat jakýkoliv typ potravy.

Problém s návrhem takového bakteroidu je v tom, že pro běžný konečný automat přijímající řetězce symbolů nad abecedou {a,b} představuje problém rozeznat, jestli přijímá řetězce, ve kterých převažují a-čka nad b-čky. Vpravdě, takový jazyk není regulární a nelze je konečným automatem rozeznat. Není ale problém naprogramovat jeden automat $\mathcal{A}$ tak, aby akceptoval jenom a-čka, a jiný automat $\mathcal{B}$ tak, aby akceptoval a-čka i b-čka. Zřejmě $\mathcal{A}$ může být pořád aktivní (není rozumné odmítat a-čka, pokud se nějaká vyskytnou). Co má smysl je vypínat $\mathcal{B}$ v případě, že je v okolí dostatek a-ček. Uvažujme proto ještě další typ automatu – automat $\mathcal{R}$ zvaný receptor. Tento automat nedělá nic jiného než že „zkoumá“ objekt na svém vstupu a pokud je to objekt a, tak $\mathcal{R}$ „vypne“ automat $\mathcal{B}$ tím způsobem, že přepne počáteční stav automatu $\mathcal{B}$ na pasivní. Pokud je na vstupu b, aktivizuje automat $\mathcal{B}$. Poté $\mathcal{R}$ objekt „pustí“ bez toho, že by ho nějak změnil, a zkoumá další náhodný objekt plující kolem.

Předpokládejme, že kódy automatů $\mathcal{A}$, $\mathcal{B}$, a $\mathcal{R}$ jsou součástí genetického kódu (řídícího programu) bakteroidu $\mathcal{S}$ a že kromě toho jsou zde přítomny i kódy pro kopírování programu, výrobu deštiček a součástí genetického kódu podobně, jako tomu bylo v případě protobakteroidu $\mathcal{P}$. Automaty $\mathcal{A}$ a $\mathcal{B}$ budeme nazývat póry, anebo kanály a budeme předpokládat, že společně s receptory $\mathcal{R}$ jsou umístněné v membráně. Jak vzniknou a jak se dostanou do membrány? Zde opět budeme spoléhat na samoorganizaci – $\mathcal{S}$ bude obsahovat automaty na výroby příslušných součástí, které se dále v membráně samoorganizací uspořádají do tvaru pórů a receptorů. Je to pouze jakýsi jiný typ deštiček, ze kterých byla konstruována původní membrána. Počet pórů a receptorů je dán množstvím živin (objektů typu A a B), které bakteroid „pozře“ a případně velikostí povrchu membrány. To vše jsou veličiny, které výpočetní mechanizmus bakteroidu nesleduje (a ani nemůže sledovat, protože je to konečný mechanizmus a nechceme, aby jeho velikost závisela na hodnotě těchto veličin).

Výsledný systém, bakteroid $\mathcal{S}$, funguje následujícím způsobem. Póry typu A procházejí selektivně pouze objekty typu A, pokud se takové v blízkosti póru vyskytnou. Receptory asynchronně zkoumají objekty na svém vstupu a aktivují či deaktivují kanály typu B. Je důležité si uvědomit, že každý receptor zavře/otevře všechny kanály a že receptory pracují asynchronně. Takže, pokud je v okolí protobakteroidu v daném časovém intervalu přítomna většina a-ček, tak většina receptorů zavře kanály typu B a tyto zůstanou

prakticky po celý interval uzavřeny, protože ojedinělé signály k jejich otevření (od receptorů, které „viděly“ b-čka) jsou okamžitě „přepsány“ signály pro zavření kanálů. V opačném případě převáží signály k otevření kanálů typu B a bakteroid „akceptuje“ potravu obou typů. Nemá cenu se zabývat podrobným „vyladěním“ celého systému – kdy nastane přepnutí z jednoho režimu do druhého – to zřejmě závisí od mnoha parametrů. Důležité ale je, že o to vše se postará evoluce, ve které zvítězí protobakterie s nejlépe fungující selektivní strategií. Všimneme si, že bakteroid $\mathcal{S}$ vykazuje známky kognitivního chování – rozlišuje mezi dobrým a nedobrým.

Právě popsaný selektivní mechanismus chování bakteroidu je základem mnoha dalších typů chování, která pozorujeme u skutečných bakterií (např. u *Escherichia coli*). Uvažujme např. bakteroid $\mathcal{M}$ opatřený flagelárním motorem. Tento motor lze zapnout tak, aby způsobil pohyb bakteroidu směrem vpřed, anebo tak, aby způsobil náhodné natočení směru plavby bakteroidu v třírozměrném moři objektů. Konstrukci stroje opět přenecháme samoorganizaci, bakteroid musí pro něj připravit pouze vhodné součásti. Zájemce o konstrukci a evoluci takových motorů ze sekrečních pórů odkazujeme např. na práci [9]. Pro naše další úvahy bude důležitý pouze předpoklad, že motor je vlastně elektromotorem, který čerpá energii z rozdílů elektrických potenciálů vně a uvnitř membrány, tj. spotřeba motoru nesouvisí s přijímanou potravou. Platí ale, že množství potravy zkonzumované bakteroidem roste s délkou dráhy, kterou bakteroid v prostředí bohatém na potravu překoná. Budeme tedy chtít, aby se $\mathcal{M}$ „pásl“ na místech s dostatkem stravy co nejdéle a v případě, že z této oblasti vyjede, aby se do ní opět vrátil.

Uvažujme proto podobný selektivní mechanizmus jako v případě bakteroidu $\mathcal{S}$, pouze s tím rozdílem, že strava typu A je „dobrá“, strava typu B „nechutná“. Naprogramujme řízení $\mathcal{M}$ takto: pokud receptory hlásí „dobrá strava“, nechejme bakteroid „pást se“ a současně plout jedním směrem. Pokud bakteroid vyjede ze zóny dobré potravy, což jeho receptory zjistí, přepne na okamžik svůj motor do druhého režimu, co způsobí jeho náhodné natočení v prostoru a vzápětí přepne opět, na krátkou dobu, na „plnou parou vpřed“. Pokud se dostane opět do zóny dobré stravy, tak pokračuje v plavbě daným směrem, jinak se opět pokusí náhodně natočit a krátkým posunem „najít“ oblast dobré potravy, atd. Zbývá pouze vyřešit problém, jak bakteroid pozná, že uplynul „okamžik“ anebo „krátká doba“. Za tím účelem využívá jakési „minutky“ – specializované automaty, jejichž činnost trvá potřebnou dobu. Příslušný automat nejprve „vypočítá“ tj. zkonstruuje nějaký objekt a vzápětí jiný automat ho začne „rozebírat“. Velikost objektů je volena tak, aby celý proces konstrukce a dekonstrukce trval potřebnou dobu.

Samozřejmě, v rámci konečných automatů není na místě používat čítače, proto je velikost objektů, které minutkový automat konstruuje, daná externě – velikostí prostoru, který objekt vyplní. Toto řešení, jakkoliv uměle vypadá, odpovídá do jisté míry skutečnosti (viz. např. [7]). Pro zajímavost, mechanizmus zapínaní a vypínaní flagelárního motoru u bakterie *Escherichia coli* pracuje přibližně takto. Na motoru je jakýsi přepínač z jednoho režimu do druhého, který lze v dané poloze fixovat tím způsobem, že se „překryje“ jistým množstvím látky, které je vymezené, zdá se, ohraničeným prostorem jistého objemu v okolí přepínače. Tato látka zamezí přístupu signálů k přepnutí přepínače. Látka se však postupně rozkládá a po nějaké době (řádově vteřiny až minuty) je opět přepínač volný. Právě popsané chování bakteroidu, zvané chemotaxe, zabezpečí, že bakteroid nevyjíždí často z oblasti dobré potravy a pokud se tak stane, má šanci se tam opět vrátit. Navenek se takové chování jeví jako inteligentní, avšak, pokud známe

mechanizmus v pozadí tohoto jevu – je to inteligence? Anebo opačně – není v pozadí jakékoliv domnělé inteligence mechanizmus, který neznáme, a proto se nám jeví příslušné chování jako inteligentní?

Abychom si učinili představu o složitosti reálných bakterií, tyto se skládají z desítek miliónů biologických molekul, mají v průměru kolem 2000 genů, v jednodušších případech několik stovek genů, převyšující 300 genů. *Escherichia coli* má 4288 genů, používá asi 20 různých „přenosových tras“ (signal pathways) pro přenos informací uvnitř buňky. Dále má na výběr cca 30 typů chemoreceptorů, které nejsou vždy všechny přítomny; bakterie je ovšem může dle potřeby „nechat narůst“. V průměru se na povrchu membrány nachází přibližně 5 typů chemoreceptorů v celkovém množství kolem 1500 [10].

7 Závěr

V článku jsme navrhli abstraktní ztělesněný hybridní systém – bakteroid, který spojuje výpočetní mechanizmy s mechanizmy samoorganizace. Pokud je nám známo, je to první systém svého druhu v oblasti ALife, o kterém lze formálně dokázat, že ve vhodně navrženém vesmíru, který je ovládán zákony, které dovolují existenci a konstrukci objektů s požadovanou samoorganizační vlastností, vykazuje známky minimálního života. Pochopení a matematické modelování sebesestavovacích procesů je v současné době předmětem intenzivního výzkumu (viz např. práci [16] která navazuje na stávající práci). Popisný rámec bakteroidu umožňuje formulaci a řešení problémů (jakými je např. problém nutriční selektivity či chemotaxe za potravou), které v jiných formalizmech nelze přirozeně formulovat ani řešit. Problémů inspirovaných schopnostmi reálných bakterií existuje celá řada (viz např. [7]) a v příštím období bude jistě zajímavé pokusit se nalézt jejich řešení. Další zajímavou výzvou do budoucna je rozšíření záběru našeho modelování z jednobuněčných organizmů na mnohobuněčné. Podobné projekty mohou vrhnout nové světlo na složitost (ve výpočetním smyslu toho slova) konstrukce a epistemické práce bakteroidů a přeneseně i na syntézu *de nuovo* reálných umělých organismů. Craig Venter a Hamilton Smith začali svůj projekt, zmíněný v úvodu této práce, také počítačovými simulacemi; jejich cílem bylo zjistit, jaký genetický materiál by mohl být použitelný pro minimální život.

Poděkování: Tato práce vznikla v rámci záměru AV0Z10300504 s částečnou podporou grantu NPV I: Informační společnost 1ET100300419

Literatura

[1] Adleman, L.: *Toward a mathematical theory of self-assembly.* Tech. Rep. 00-722, Dept. of Computer Science, University of Southern California 2000.

[2] Benenson, Y., Paz-Elizur, T., Adar, R., Keinan, E., Livneh, Z., Shapiro, E.: Programmable and autonomous computing machines made of biomolecules. *Nature* **414** (2001) 430-434.

[3] Brooks, R.: Beyond Computation. *EDGE* **132**, June (2002) (http://www.edge.org/3rd_culture/ brooks_beyond/beyond_index.html).

[4] *Central dogma of molecular biology*: viz např. http://www.bartleby.com/59/21/ centraldogma.html.
[5] Hanczyc, M. M., Fukijawa, S. M., Szostak, J. W.: Experimental Models of Primitive Cellular Compartments: *Encapsulation, Growth, and Division. Science* **302** (2003) 618-622.
[6] Hutton, T. J.: Simulating Evolution's First Steps. In: *Advances in Artificial Life. Proceedings of the 7th European Conference, ECAL2003*. Banzhaf W., Christaller T., Dittrich P., Kim J.T., Ziegler J. (Eds.), Dortmund, Germany, 14-17th September 2003. LNAI 2801, Springer 2003, 51-58.
[7] Kováč, L., Nosek, J., Tomáška, Ľ.: An Overlooked Riddle of Life´s Origin: Energy Dependent Nucleic Acid Unzipping. *J. Mol. Evol.* 57 (2003) S182-S189.
[8] Lengeler. J. W., Müller, B. S., di Primio, F.: Neubewertung kognitiver Leistungen im Lichte der Fähigkeiten einzelliger Lebewesen. *Kognitionswissenschaft* **8** (2000) 160-178.
[9] Mullen, L.: Life from scratch? *Astrobiology Magazine,* 2004 (http://www.astrobio.net/ news/article319.html).
[10] Musgrave, I.: Evolution of the Bacterial Flagellum. In: *Why Intelligent Design Fails: A Scientific Critique of the Neocreationism.* Young, M., and Edis, T. (Eds.), Rutgers University Press, Piscataway, N.J. 2005, 240 p.
[11] Müller, B. S., di Primio, F., Lengeler, J. W.: Contributions of Minimal Cognition to Flexibility. In: *SCI 2001*, *Proc. of the 5th World Multi-Conference on Systemics, Cybernetics and Informatics.* N. Callaos, W. Badawy and S. Bozinovski (Eds), Volume XV, Industrial Systems: Part II, International Institute of Informatics and Systemics 2001, 93-98.
[12] Paun, G., Rozenberg, G.: A Guide to Membrane Computing. *Theoretical Computer Science* **287** (2002) 73–100.
[13] Pfeifer, R., Scheier, C.: *Understanding Intelligence*. Cambridge, MA: The MIT Press 1999.
[14] Szostak, J.W., Bartel, D.P., Luisi, P.L.: Synthesizing Life. *Nature* **409 (**2001) 389-390.
[15] Wiedermann, J., van Leeuwen, J.: The Emergent Computational Potential of Evolving Artificial Living Systems. *Ai Communications* **15** (2002) 205-216.
[16] Wiedermann, J.: Building a Bridge between Mirror Neurons and Theory of Embodied Cognition. In: *Proc. SOFSEM 2004*, P. van Emde Boas et al. (Eds.), LNCS 2932, Springer, 2004, 361-372.
[17] Wiedermann, J.: Globular Universe and Autopoietic Automata: A Framework for Artificial Life. In: *Advances in Artificial Life*. (Ed.: Capcarrere M., Freitas A.A., Bentley P.J., Johnson C.G., Timmis J.) - Berlin, LNCS Vol. 3630, Springer-Verlag 2005, 21-30.
[18] Wiedermann, J.: Computing by Self-Reproduction: Autopoietic Automata. In: *Unconventional Computation*. (Ed.: Calude C. S., Dinnen M. J., Paun G., Perez-Jimenez M. J., Rozenberg G.) Berlin, LNCS 3699, Springer-Verlag 2005, 224-236.
[19] Wiedermann, J.: 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.

[20] Winfree, E.: DNA Computing by Self-Assembly. *National Academy of Engineering Website, The Bridge* **33** (4) (2003) 31-38.
[21] Zimmer, C.: Building a Virtual Microbe, Gene by Gene by Gene. The New York Times, August 16, 2005.

Index autorov

Andoga, Rudolf, 121
Běhounek, Libor, 129
Beňušková, Ľubica, 145
Bída, Michal, 157
Brandejský, Tomáš, 243
Brom, Cyril, 157, 257
Burian, Jan, 173
Cienciala, Luděk, 369
Farkaš, Igor, 3
Grešková, Mirka, 191
Hajnal, László, 17
Havel, Ivan M., 27
Horáková, Jana, 71
Hvorecký, Juraj, 89
Ivančo, Marián, 389
Jirků, Petr, 209
Kadlec, Rudolf, 157
Kelemen, Jozef, 71
Kelemenová, Alica, 369, 399
Kopčo, Norbert, 121
Kováč, Ladislav, 415
Kůrková, Věra, 275
Kvasnička, Vladimír, 441, 475
Kvasz, Ladislav, 95
Lúčny, Andrej, 287
Lukavský, Jiří, 257
Markoš, Anton, 465
Markošová, Mária, 297
Mikulecký, Peter, 309
Náther, Peter, 297
Pospíchal, Jiří, 475
Pstružina, Karel, 111
Rybár, Ján, 223
Sarnovský, Ján, 317
Sinčák, Peter, 329
Sýkora, Libor, 465
Sýkorová, Lenka, 465
Šefránek, Ján, 233
Takáč, Martin, 339
Trebatický, Peter, 357
Wiedermann, Jiří, 497