

# Konceptuálne východiská pre model stelesnenej mysle

Igor Farkaš<sup>1</sup>

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

## 1 Úvod

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

Zodpovedať na tieto otázky v jednom príspevku je nereálne už len z toho dôvodu, že všeobecne akceptované odpovede nepoznáme. Ja sa pokúsim zúžiť a prezentovať svoj pohľad na podstatu *ľudskej* mysle a predovšetkým jej ukotvenie v materiálnom tele. Dnes už snáď málokto pochybuje o tom, že ľudská myseľ ako nehmotná entita sídli v hmotnom tele, pretože Descartov dualizmus substancii už dnes ťažko možno považovať za vedeckú alternatívu (jemnejšie formy dualizmu, napr. dualizmus vlastností, sú už prijateľnejšie). Na druhej strane, redukcionistické teórie identity mozgu-mysle alebo eliminačného materializmu tiež nie sú podľa mňa tým „pravým orechovým“, lebo úplne eliminujú a zavrhnú napr. fenomenologické vlastnosti mysle. Tu prezentovaný prístup možno zaradiť k platforme *neredukcionistického fyzikalizmu* (Davidson, 1970), lebo ten síce kladie dôraz na hmotný svet, a teda sa snaží vysvetliť kauzálne mechanizmy len v rámci neho (kauzálna uzavretosť), avšak berie do úvahy, že opis mentálnych vlastností (ich epistemologické poznanie) a ich vysvetlenie nie je redukovateľné na opis jazykom prírodných vied. V tejto súvislosti možno niekomu napadne otázka, že načo potom existuje mentálna úroveň (myseľ), keď takýto prístup ju vlastne nepotrebuje, pretože myseľ nemá kauzálny vplyv na mozog, čo sa javí byť v rozpore so subjektívnou skúsenosťou. Rozpor je podľa mňa zdanlivý; argumenty uvádzam v závere príspevku (časť 6).

Keďže vychádzam z predpokladu o kauzálnej uzavretosti fyzikálneho sveta, budem nevyhnutne hovoriť aj o mozgu (resp. celom nervovom systéme), a o procesoch v ňom, pretože tie by mohli zodpovedať za to, čo sa deje v mysli človeka, aj keď „len“

---

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

z pozície tretej osoby (3PP).<sup>2</sup> Pri vedeckom štúdiu mozgu máme k dispozícii experimentálne prístupy skúmajúce správanie človeka, moderné zobrazovacie metódy (činnosti) mozgu, ktoré umožňujú lokalizovať aktivitu mozgu pri mentálnej činnosti, a výpočtové modelovanie, ktoré umožňuje explicitne skúmať neurálne mechanizmy na zvolenej úrovni abstrakcie. Pozícia prvej osoby (1PP), typická skôr pre fenomenológiu (Havel, 2007) je však, myslím si, tiež dôležitá, pretože nie je redukovateľná na opis z 3PP. Vyžaduje si alternatívne prístupy skúmania, ktoré by mali byť spoľahlivejšie ako introspekcia zavedená v 19. storočí (Chalmers, 1999). Niektorí filozofi vidia väčšie prepojenie oboch perspektív. Napríklad Searle (1999), ktorý chápe mentálne stavy ako čisto fyzikálny fenomén, pretože tie sú produktom mozgu, argumentuje, že subjektivnosť ontológie mentálnych stavov mysle nie je prekážkou ich epistemickej objektívnosti poznávania (príklad: napriek tomu, že vnímanie bolesti je číro subjektívne, môžeme mať objektívnu vedu skúmania bolesti).

## 2 Exkluzívne fundamenty ľudskej mysle

Dôsledkom toho, že naša myseľ neexistuje „v éteri“, ale je ukotvená v materiálnom svete je fakt, že naše mentálne stavy dokážeme skrývať pred vonkajším svetom len čiastočne, pretože naše telo ich *prezrádza*. Niektoré naše vnútorné stavy (hlavne tie, ktoré sú výrazne citovo podfarbené) do istej miery odkrývame (zámerne alebo nechtiac) smerom k iným subjektom, a to svojimi gestami, náznakmi, či správaním všeobecne.<sup>3</sup> Na základe týchto indikátorov každý z nás dokáže čiastočne čítať mentálne stavy iného subjektu, čo potvrdzuje priame prepojenie nášho vnútra a vonkajších prejavov nášho tela. Schopnosti porozumieť vnútru iného subjektu na základe týchto behaviorálnych indícií sa odborne hovorí *teória mysle*, ktorou disponujeme pravdepodobne ako jediný živočíšny druh. Veľký význam pre vysvetlenie jej fungovania bolo objavenie tzv. zrkadliacich neurónov (v premotorickej frontálnej kôre), ktoré sa pravdepodobne podieľajú na neurálnom ukotvení teórie mysle (Rizzolatti a spol., 1996). Tieto neuróny sú aktivované nielen, keď človek sám vykonáva nejakú činnosť, ale aj keď pozoruje iného človeka pri tej činnosti. Vtedy zrkadliace neuróny reagujú na pozorovanú činnosť a umožňujú takto pochopiť správanie iného človeka, vcítiť sa do jeho perspektívy.

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

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

<sup>3</sup> Samozrejme, nejaký chladnokrvný jedinec, či skvelý herec sa dokáže navonok tváriť indiferentne.

pohyby (ručné práce, tanec, hudba, šport), čo nedokáže žiaden iný živočích. Všetky tieto činnosti v sebe skrývajú schopnosť mozgu spracovávať sekvencie dát, čo vystihuje hypotéza o „sekvenčnej kôre“ (Calvin & Ojemann, 1994, kap. 16).

Naše exkluzívne ľudské schopnosti sú produktom mozgu a súčasne prejavmi mysle. Utvrdzujú z nás v presvedčení, že aj napriek tomu, že ja nemám ako zistiť, že aj Vy máte myseľ, predpokladám to na základe podobnosti (či kvalitatívnej rovnakosti v rámci ľudského druhu). Prirodzenou ľudskou snahou je pochopiť podstatu našich mentálnych prejavov, a snahou niektorých je aj reprodukovať ich (alebo aspoň niektoré z nich) v nebiologickom svete. Napríklad, teória mysle sa stala predmetom záujmu aj pri návrhu inteligentných agentov (Wiedermann, 2003). Nato, aby sme pochopili mechanizmy mentálnych procesov, sú veľmi dobrým pomocníkom výpočtové modely. Čím exaktnejší model, tým lepšie. Na rozdiel od menej detailných teoretických modelov typu „box-and-arrow“ sprevádzaných verbálne formulovanou teóriou, sú exaktné výpočtové modely vďaka svojej matematickej explicitnosti jednoznačné, čo sa týka ponúkaného vysvetlenia mechanizmov, a navyše ponúkajú testovateľné predikcie.

### 3 Výpočtové prístupy k modelovaniu mysle

Naše vlastné mentálne stavy (presvedčenia, ciele, nálady, atď.) a procesy (myslenie) sú to jediné, k čomu máme prístup. Aj keď zrejme drvivá väčšina z nás bude súhlasiť, že ľudská myseľ je stelesnená (resp. vtelená) v hmotnej realite, nie každý bude asi súhlasiť, že pri snahe pochopiť fungovanie mysle musíme nutne hovoriť o mozgu a neurálnych procesoch. Tento pohľad o existencii potenciálne nezávislej, vyššej úrovne ontologickej reality existoval už v antickej filozofii mysle, a v polovici minulého storočia bol výrazne posilnený nástupom počítačov (analógia mysle-mozgu so softvérom-hardvérom počítača), ktoré významne posilnili ovplyvnili názorové spektrum v kontexte odvekého problému vzťahu mozgu a mysle (Farkaš, 2007). Inými slovami, dá sa argumentovať, že napriek tomu, že mozog je sídlom našich mentálnych stavov, na ich opis nemusíme používať neurálne koncepty, ale mentálne koncepty, a ukotvenie modelu mysle je len vecou implementácie. Tento pohľad bol posilnený aj zavedením troch nezávislých úrovní abstrakcie – výpočtovej, algoritmickej a implementačnej<sup>4</sup> (Marr, 1982), čo bolo motivované práve počítačovou metaforou mozgu. Marrov dôraz na prvú a druhú úroveň bol konzistentný s počiatočnými trendmi v umelej inteligencii, kognitívnej psychológii a kognitívnej vedy, ktoré boli postavené na predpoklade, že na implementačnej úrovni abstrakcie nezáleží. Používanie počítačovej metafory mozgu je síce možno nápomocné ako veľmi hrubá predstava o fungovaní mysle, no z istého pohľadu môže byť dosť zavádzajúce. Že počítač je počítačom, a teda zariadením realizujúcim výpočty, je zřejmé, no či aj mozog je počítač, také zřejmé nie je. Prečo?

#### 3.1 Robí ľudská myseľ-mozog výpočty?

Výpočet (nejaký algoritmus) možno realizovať rôznymi spôsobmi, počnúc perom a papierom, a končiac implementáciou v počítači. Pohľad na mentálne procesy pomocou zavedenia formálneho opisu má nesporne svoje výhody, lebo umožňuje navrhovať a študovať mechanizmy, v snahe vysvetliť správanie ľudí (porovnaním s predikciami modelu). Ak sú mentálne procesy výpočtom, potom sa musíme opýtať: akým typom

---

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

výpočtu, s čím, s akými entitami (reprezentáciami)? Ak ide o výpočty, nemožno takéto výpočty realizovať na inom hardvéri než je mozog? Môže mať napríklad stroj mentálne stavy? Otázka, či mozog-počítač by bol výpočtovým strojom je triviálna, pretože počítač bol navrhnutý ako výpočtový stroj, ktorý komunikuje s prostredím pomocou svojich vstupno-výstupných zariadení, ktoré fungujú ako prevodníky – prevádzajú aktivitu z jedného média do druhého (Dennett, 1996). Spracovanie informácie vnútri počítača je teda „homogénne“ (ide o prenos elektrických aktivít). Okrem toho, v počítači je jednoznačne určená implementačná úroveň, na ktorej výpočty prebiehajú (logické hradlá). V prípade mozgu však implementačná úroveň vôbec nie je jednoznačná, pretože mozog nikto nezostrojil za účelom počítania. Ten sa evolučne vyvinul tak, aby umožnil organizmu efektívne fungovať v dynamickom prostredí (Churchland, Sejnowski, 1992). Na akej úrovni potenciálne prebiehajú výpočty v mozgu? Najpriateľnejšou hypotézou sa javí úroveň neurónov, hoci existujú neurofyziologické poznatky o tom, že nižšia (molekulárna) úroveň fyzikálnej reality ovplyvňuje vyššiu úroveň, t.j. činnosť samotných neurónov (Bell, 1999). Podobne, rôzne (pomalšie) hormonálne a neuromodulátorové systémy ovplyvňujú správanie neurónu (Kaczmarek & Levitan, 1987).<sup>5</sup> Okrem toho, spracovanie informácie v mozgu má elektricko-chemický charakter<sup>6</sup> a de facto sa opiera o fyzikálne procesy, ktoré dokážu fungovať v časovej škále optimalizovanej pre existenciu človeka v dynamickom prostredí. Taktiež na rozdiel od počítača, kde môžeme izolovať hardvér od softvéru, a kde hardvér dvoch počítačov toho istého typu je rovnaký, každý mozog je jedinečný, a tak konkrétne mentálne prejavy sú dosahované individuálnymi konfiguráciami fyzikálnych stavov. Na tom, z čoho je skonštruovaný mozog, teda možno záleží.

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

### 3.2 Vlastnosti ideálneho modelu mysle

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

- A) *robustné správanie* – postupná degradácia fungovanie v prípade šumu či čiastočných porúch systému
- B) *adaptívnosť* – schopnosť učenia, prispôsobovania sa meniacim sa vonkajším podmienkam (prostredia)
- C) *reprezentácia znalostí* – model musí mať vnútorne reprezentované nadobudnuté poznatky, ktoré vie využívať, a ktoré neustále aktualizuje (počas učenia)
- D) *empirická prijateľnosť predikcií modelu* – správanie modelu musí byť v zhode s prejavmi správania človeka. Mám na mysli modelovanie procesov:
  - 1. *nižšej kognície*: vnímanie, rozpoznávanie obrazcov, konanie, senzomotorická koordinácia, a i.
  - 2. *vyššej kognície*: usudzovanie, plánovanie, riešenie problémov a i.

---

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

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

3. *všetkých ostatných procesov*, ktoré nevieme jednoznačne zaradiť, ako napr. používanie jazyka,<sup>7</sup> priestorová kognícia, imaginácia, pamäť a i.
- E) *vysvetlenie recipročnej kauzality* medzi myslou a mozgom-telom – pretože nás k tomu nabáda vlastná skúsenosť,
- F) *vysvetlenie subjektívnosti* – do akej miery, a či vôbec, vie model vysvetliť, prečo vôbec existuje IPP.

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

## 4 Hlavné paradigmy v kognitívnej vede

### 4.1 Symbolová paradigma

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

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

---

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

Predstavujú v súčasnosti najschodnejšiu cestu ako modelovať myšlienkové pochody človeka sprevádzajúce úlohy vyššej kognície.

Negatívom symbolových systémov je, že celú znalosť (bázu znalostí i pravidiel) do systému musí vkladať dizajnér, znalý domény. Učenie hrá minoritnú úlohu. Taktiež, kde vznikajú problémy, sú úlohy vyžadujúce kognitívne procesy na nižšej úrovni, ktoré človek takisto poľahky zvláda. Príkladom môže byť vizuálne rozpoznanie nejakého objektu, alebo hlasu človeka. Je zrejmé, že v týchto prípadoch sa človek opiera o vizuálne či akustické charakteristiky vstupu, a na základe ich spracovania dospeje k stavu rozpoznania, spojeného s nejakým mentálnym stavom (*áno, je to on!*). Symbolové systémy sa takéto úlohy tiež pokúšali riešiť, a to prepísaním vstupných dát do symbolovej formy a ich následným spracovaním. Takéto snahy sa však stretli s neúspechom, a to nielen kvôli vlastnostiam dát (vysoká rozmernosť, množstvo, variabilnosť, redundancia), ale hlavne kvôli zložitosti navrhnuť veľké množstvo pravidiel, ktoré by zabezpečili aspoň prijateľné fungovanie systému, a ktorých dizajn leží na ramenách programátora. Tu sa priamo ponúka prístup „zdola,“ čo je charakteristickou črtou konekcionalistického prístupu.

## 4.2 Konekcionalistická paradigma

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

Z pohľadu symbolovo-subsymbolovej dichotómie môžeme rozlíšiť dva typy konekcionalistických modelov. *Lokalistické* modely sú podobné symbolovým modelom v tom zmysle, že každý umelý neurón reprezentuje (symbolizuje) nejakú entitu. Avšak v *distribuovaných* modeloch sú jednotlivé entity reprezentované množinou neurónov (tzv. populačné kódovanie), čo má niekoľko výhod: umožňuje *lepšie reprezentovať sémantické koncepty* a jemné rozdiely medzi nimi a vnášajú do systému *robustnosť*. Vďaka robustnosti má systém vlastnosť *postupnej degradácie*, pretože i v prípade šumu či lokálnych porúch si zachová istú funkčnosť, ktorá bude závisieť od miery poškodenia. Na druhej strane, problémom modelov neurónových sietí práve kvôli distribuovanosti informácie je, ako súčasne reprezentovať viacero konceptov tými istými neurónmi (tzv. problém viazania).<sup>8</sup> Druhou ťažkosťou s tým súvisiacou je reprezentácia štruktúrovaných entít (dáta so stromovou, či grafovou štruktúrou),

---

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

typických pre vyššiu kogníciu.

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

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

### 4.3 Dynamická paradigma

Niektorí autori prišli s myšlienkou rozšíriť koncept mentálnych stavov (kognície) vo vzťahu k okolitému prostrediu, pretože viacero aktivít, ktoré človek vykonáva, sa dá chápať ako výsledok permanentnej interakcie s prostredím (Thelen & Smith, 1994; van Gelder & Port, 1995). Kognitívne (mentálne) procesy teda nie sú chápané izolovane od prostredia, ale v bezprostrednej previazanosti s ním (coupling). Kognitívne procesy s tým súvisiace sú mentálnou aktivitou v čase, modelovateľnou pomocou systému nelineárnych diferenciálnych rovníc. Táto paradigma, postavená na teórii dynamických systémov predpokladá existenciu dvoch kľúčových, vzájomne prepojených typov premenných. Jedným sú tzv. *kolektívne premenné*, ktoré zahŕňajú vzťahy medzi interagujúcimi zložkami dynamického systému, a vysvetľujú správanie systému. Druhým typom sú *riadiace parametre*, ktorých kvantitatívne zmeny môžu spôsobovať kvalitatívnymi zmenami v správaní systému (tzv. fázové prechody). Príkladom takéhoto systému môže byť pohyb koňa, ktorého rytmus pohybu (vyjadreného periodickými priebehmi kolektívnych premenných) závisí od rýchlosti pohybu (riadiaci parameter). Koň sa môže pohybovať v každom čase len v jednom rytme, (fázové) prechody medzi nimi nemusia nutne nastávať pri tých istých rýchlostiach, a závisia aj od stavu prostredia (napr. hrbolatá cesta). Iným príkladom môže byť učenie sa chodiť u batoliat,

---

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

kde taktiež dochádza k bezprostrednej interakcii s najbližším okolím. Dynamický opis je teda postavený na premenných, ktoré treba identifikovať. V prípade bezprostrednej senzomotorickej interakcie s prostredím sa to zrejme dá. Dynamický pohľad na spoznávanie vonkajšieho sveta je úzko spojený so zjednávacím (enactic) prístupom (Varela a spol., 1991), resp. s pohľadom známym ako stelesnená kognícia, v ktorom sa dôraz z vnútorných reprezentácií (vopred daného) vonkajšieho sveta presúva na *vnímanie a jednanie vo svete*, ktorý sa takto spoluvytvára. Tým vzniká prostredníctvom spätnej väzby kruhový proces, pretože zjednávaný svet je prostredím pre zjednávanie, ale zjednávanie je súčasne predpokladom zjednávaného sveta. V rámci tohto pohľadu si každý subjekt (kognitívny systém) zjedná svoj svet sám v procese dynamickej interakcie, a ten bude jedinečný.

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

**Tabuľka 1.** Hlavné paradigmy kognitívnej vedy.

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

#### 4.4 Podobnosti a rozdiely medzi paradigmami

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

Silné a slabé stránky oboch prístupov sa dajú dobre ilustrovať v kontexte modelovania jazykovej schopnosti človeka, ktorá sama osebe obsahuje podprocesy na rôznych úrovniach kognície (od fonológie po pragmatiku). Je pozoruhodné, že jazykový prejav človeka sa ukazuje vo väčšom súlade s prednosťami a ťažkosťami konekcionistických modelov než so symbolovými systémami (Rohde & Plaut, 2003;

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



Farkaš, 2005). Napríklad, človek poľahky chápe jemné sémantické rozdiely v slovách či na úrovni viet, no má problémy so spracovaním zložitých rekurzívnych viet. V symbolovom systéme to býva typicky naopak. Psychologické dáta sú v pomerne dobrom kvalitatívnom a často i kvantitatívnom súlade s predikciami konekcionistických modelov (napr. Rohde, 2002; Mayberry, 2003; O'Reilly & Munakata, 2000). Problémom ostáva efektívne učenie týchto zložitých modelov, ktoré bráni ich zväčšeniu (slovníka a komplexnosti použitej gramatiky).

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

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

#### 4.5 Hypotéza o symbolových operáciách pri pohľade „zdola“

Domnievam sa, že medzi symbolovými a subsymbolovými systémami existuje istá asymetria. Otázka vysvetlenia procesov nižšej kognície je prakticky neschodná pre symbolové systémy ako som argumentoval v závere časti 4.1. Z opačného pohľadu, vysvetlenie mechanizmov sprevádzajúcich vyššiu kogníciu pomocou nejakého adaptívneho kolektívneho systému je možno len otázkou času kvôli ťažkostiam tohto prístupu. Tu vzniká otázka, ktorú Clark (2001, s. 135) označil termínom *kognitívna inkrementálnosť*: „Aký je vzťah medzi stratégiami používanými pri základných

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

problémoch vnímania a konania, a tými, ktoré používame na riešenie abstraktnejších problémov na vyššej úrovni?“ Inými slovami, je medzi nimi kvalitatívny rozdiel (z hľadiska používaných mechanizmov), alebo sú ukotvené v mozgu rovnakým spôsobom na nejakej konekcionisticky relevantnej úrovni abstrakcie? Mohol počas evolúcie nastať nejaký kvalitatívny skok vo vývoji mozgu človeka, ktorý by súvisel s našimi schopnosťami vyššej kognície? Je známe, že evolúcia využíva staré, osvedčené mechanizmy aj pri návrhu nových systémov. Napríklad, biologické mechanizmy plasticity a učenia sú evolučne staré (človek ich má podobné, a v niektorých aspektoch rovnaké, ako nižšie živočíchy), a v rámci ľudského mozgu vyzerajú byť univerzálne, či už hovoríme o vizuálnej kôre alebo frontálnej kôre (Jedlička a spol., 2002).<sup>12</sup>

Dokážeme ponúknuť hypotézu, ako pomocou kolektívneho systému vysvetliť schopnosť človeka spracovávať symbolové dáta? Vo svete okolo nás existujú dáta subsymbolového (numerického) charakteru (obrazy, zvuky, atď.) i symbolového charakteru (matematické vzorce, schémy a pod.). Človek nepochybne disponuje schopnosťou spracovávať aj symbolové dáta (aj keď ich de facto prijíma cez senzorické vstupy v subsymbolovej forme). Je však pozoruhodné, že pri aplikácií rôznych transformácií na symboloch (napr. pri matematických úpravách) si človek pomáha *externou pamäťou*, napr. zapisovaním medzivýsledkov na papier pri násobení čísel. Je to zrejme preto, aby bol schopný mentálne realizovať každý krok – symbolovú operáciu ako vektorovú transformáciu (distribuuwanej reprezentácie, t.j. vzorca aktivít), lebo takéto jedнокrokové transformácie podporuje ľudská kognitívna architektúra.<sup>13</sup> Druhým dôvodom je naša obmedzená kapacita pracovnej pamäti. Povedané slovami Hutchinsa (1995), „výpočtová sila ľudského prístupu nezávisí primárne na našom vnútornom informačnom zariadení (mozgu), ale na spolupráci s externou súčasťou technológie, ktorá nám umožňuje realizovať úlohy, v ktorých sme dobrí: rozpoznávanie vzorov, modelovanie jednoduchej dynamiky a manipulácia s objektmi v priestore“. Ak by bola počítačová metafora pravdivá, prečo potom nevieme robiť celé sekvencie operácií na symboloch v hlave?

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

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

---

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

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

#### 4.6 Kauzalita vo svetle paradigiem

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

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

(a) *na jednotlivých úrovniach* – t.j. v kauzálnych doménach (Havel, 2001), v rámci ktorých môžeme hľadať príčinné vzťahy medzi javmi. Kauzálne chápaný vzťah v jednej doméne (vidím hada, preto znehybniem), môže byť vysvetlený sprievodne v inej doméne (neuróny vo vizuálnej kôre registrujú hada, čo spôsobí vhodnú konfiguráciu aktivít neurónov v motorickej kôre, ktoré blokujú svalové pohyby). Toto môžeme nazvať *kauzálnym paralelizmom* (Havel, 2001, s. 46).

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

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

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

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

Konekcionistické modely sú zase kvôli svojej distribuovanosti spájané s javom, ktorý Ramsey a spol. (1990) nazvali *totálny kauzálny holizmus* (s. 181). Ten podľa

<sup>14</sup> Pri radikálnejšej verzii dynamickej paradigmy sa recipročná kauzalita rozplýva v podobe kontinuálnej interakcie medzi mozgom, telom a prostredím v reálnom čase.

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

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

#### 4.7 Systematickosť mysle

Fenomén systematickosti ľudskej mysle sa taktiež priamo dotýka spomínaných paradigiem, a najmä pre konekcionizmus predstavuje stále veľkú výzvu (Aizawa, 2003). Možno ho najlepšie vysvetliť pomocou analógie so schopnosťou človeka systematicky rozumieť jazyku.<sup>18</sup> Ak niekto vie povedať, že *Janko ľúbi Aničku*, bez problémov vie vytvoriť aj vetu *Anička ľúbi Janka*, pretože vie, že jazyk sa skladá z častí a vie, ako ich skladať pri tvorbe viet. Systematickosť v jazyku teda implikuje *produktívnosť* (schopnosť generovať teoreticky neobmedzené množstvo viet) a *kompozičnosť* (existencia vnútornej štruktúry viet).<sup>19</sup> V symbolových modeloch je problém systematickosti automaticky vyriešený, pretože reprezentácie a syntaktické operácie nad nimi (podľa pravidiel) priamo umožňujú produktívnosť symbolového systému. V konekcionistických systémoch takéto vlastnosti na prvý pohľad vôbec nie sú zrejmé, a podľa niektorých kritikov ani principiálne možné. Fodor a Pylyshyn (1988) vo svojom vplyvnom kritickom článku sformulovali svoj argument zhruba nasledovne: „Ľudská myseľ je systematická, a preto jej interné reprezentácie sú štruktúrované. Keďže v konekcionistických systémoch neexistujú štruktúrované reprezentácie, tie (systémy) nie sú dobrými modelmi ľudskej mysle.“ Odozva na túto kritiku prišla v podobe niekoľkých modelov, v ktorých autori poukazovali na existenciu systematickosti distribuovaných reprezentácií v rôznych (rekurzívnych) modeloch neurónových sietí (napr. pozri prehľad vo Farkaš, 2005, časť 3.3). Pozoruhodným poznatkom je to, že distribuované konekcionistické reprezentácie majú tzv. *funkčnú* kompozičnosť (van Gelder, 1990) bez nutnosti *konkatenatívnej* kompozičnosti, typickej pre symbolové systémy, ktorá vzniká zret'azením reprezentácií jednotlivých elementov (konštituentov).

---

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

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

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

<sup>19</sup> Kompozičnosť predpokladá možnosť syntakticky preusporiadať konštituenty reprezentácie.

Podstatné je, že funkčná kompozičnosť umožňuje realizovať štruktúralne závislé transformácie bez nutnosti extrakcie elementov (z vektorov) a ich preusporiadania (napr. zámena subjektu a objektu).

#### 4.8 Propozície či mentálne obrazy?

Systematickosť sa priamo týka aj jazyka mysle (mentálčiny). Ak niečo také existuje (ontologická otázka), otázkou je, akú to má povahu – či jazykovú (propozičnú) alebo obrazovú. Propozičná mentálčina (Pylyshyn, 1984) má blízko k formálnym jazykom opisu a môžeme pomocou nej vyjadriť prakticky čokoľvek. Propozície<sup>20</sup> sa dajú vnárať, kombinovať, môžeme zaviesť ľubovoľné typy propozícií (s rôznymi hlavami a typmi argumentov). Napríklad to, že *Janko má rád Marienku, ktorá venčí psa* by sme vyjadrili ako (*Love John (Is Mary (Walk Mary Dog))*). Propozície majú symbolový charakter, umožňujú opísať aj abstraktné entity a v podstate ľubovoľné vzťahy medzi nimi. Zisťovanie pravdivosti propozície (v nejakom mentálnom svete) je záležitosťou interpretácie symbolov a ich vzťahov. Keďže symboly sú amodálne (v arbitrárnom vzťahu s vonkajším svetom), vystupuje tu problém ukotvenia symbolov (Harnad, 1990). Ten sa snažili autori riešiť rôznymi spôsobmi (pozri prehľad vo Vavřečka, 2006).

Na druhej strane, mentálne obrazy opisujú entity sveta (objekty, udalosti) pomocou schém, či (hlavne) vizuálnych obrazov, ktoré zdieľajú charakteristiky s perceptuálnymi systémami (Barsalou, 1999). Keďže sú spriahnuté s percepciou, majú analógový, a teda subsymbolový charakter (Kosslyn 1994). S mentálnymi obrazmi je konzistentná teória o tzv. situačných modeloch (Zwaan & Radvansky, 1998), ktoré si človek vytvára v mysli pri spracovaní symbolového vstupu (jazyka). Inými slovami, aby človek pochopil nejakú vetu, skonštruuje si v mysli obraz situácie a keď tá je konzistentná, človek pochopil význam. V súčasnosti v literatúre neexistuje konsenzus o tom, ktorý typ reprezentácie – propozičný či obrazový, je kognitívne relevantný, pretože sa zdá, že existujú empirické poznatky v prospech oboch (Sternberg, 2003). Neexistencia konsenzu pramení aj z toho, že konkrétne empirické poznatky môžu byť interpretovateľné z oboch pohľadov. Z nášho pohľadu majú mentálne obrazy oveľa bližšie k distribuovaným reprezentáciám v rámci kolektívneho systému než propozície.

### 5 Kritériá pre vedomé stavy z pohľadu neurovedy

Pri formálnom prístupe k uchopeniu mysle nemáme principiálne možnosť vysvetliť aspekt subjektívnosti (1PP) mentálnych stavov (podmienka F). Máme však možnosť skúmať, aké kritériá musia spĺňať mentálne stavy z pohľadu 3PP. Ak máme na mysli neurálne procesy, hovoríme o tzv. neurálnych korelátoch mentálnych stavov (Metzinger, 2000). Ak pohľad zúžime na výpočtové aspekty, hovoríme o výpočtových korelátoch (Cleeremans, 2005). Zatiaľ čo neurálne koreláty sú viazané na substrát, v ktorom sú ukotvené, výpočtové koreláty (ak sa ukáže, že existujú) potenciálne umožňujú realizovateľnosť mentálnych procesov u človeka aj na inom substráte (funkcionalistický pohľad). A možno má zmysel skúmať oba typy korelátov a vzťah medzi nimi (Cleeremans, 2005). Mentálne stavy z pohľadu mysle ako kolektívneho systému (3PP) chápeme ako distribuované vzorce aktivít (na nejakej úrovni abstrakcie).

---

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

Niektoré z týchto vzorcov súvisia<sup>21</sup> s *vedomím* (čo je evidované odpovedajúcim stavom v IPP), niektoré nie. V mozgu prebieha stále množstvo procesov, ku ktorým nemáme prístup, sú nevedomé (Dehaene & Naccache, 2001). Vedomie teda chápeme ako špecifickú podtriedu možných mentálnych stavov, ktorej charakteristiky by sme radi odhalili. Vedomé mentálne stavy však nie sú nejakou homogénnou podtriedou, o čom nás presvedčajú každodenné skúsenosti. V odbornej literatúre (napr. Edelman, 1989, pozri aj Havel, 2001, časť 1.8) sa dočítame o typickom delení stavov vedomia na (1) primárne vedomie (senzorické), ktoré vzniká pri vnímaní nejakého pozorovaného objektu, (2) reflexívne (sekundárne), vďaka ktorému dokážem reflektovať vnímané podnety vo svojom „ja“ (som to ja, kto vníma, autobiografia), a (3) vyššie (rozšírené) vedomie, ktoré mi umožňuje zahrnúť i časovú dimenziu do vnútorného modelu sveta i seba (Edelman, 1989).<sup>22</sup> Je možné, že reflexívne vedomie je vyhradené len človeku a niektorým cicavcom (napr. ľudopopy sa v zrkadle spoznajú), a vyššie vedomie je výhradne vlastnosťou ľudskej mysle (Kráľ & Hulín, 2001). Možno schopnosť mysle človeka reflektovať seba samého v časovom kontexte súvisí s jeho jazykovou schopnosťou (Beňušková, 2005).

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

## 5.1 Neurálne koreláty mentálnych stavov vedomia

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

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

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

---

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

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

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

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

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

S týmito základnými vlastnosťami súvisia aj ďalšie podmienky pre vedomé stavy (Seth a spol., 2005), ako napr.: široký rozsah ich „obsahov“ (percepčne orientované stavy, pocity, myšlienkové procesy, atď.), limitovaná kapacita (človek si v danom čase môže byť vedomý len jednej „veci“), trvanie (rádovo zopár stoviek milisekúnd trvá, aby vznikol vedomý vnem), stabilita obsahu (napriek potenciálnej variabilnosti podnetu), uľahčenie učenia, nutnosť vedomia pri niektorých úlohách (napr. rozhodovanie sa) a ďalšie. Zaujímavou vlastnosťou primárneho vedomia je senzorické viazanie<sup>24</sup>, ktoré sa vysvetľuje dvoma triedami hypotéz. Na jednej strane stoja hypotézy o synchronných osciláciách vznikajúcich v dynamicky sa tvoriacich neurálnych skupinách (napr. Crick & Koch, 1998, Singer, 1999). Na druhej strane stoja hypotézy o senzorickom viazaní prostredníctvom pozornosťných procesov, exekutívnych mechanizmov a neurálnych máp (napr. Shafritz a spol., 2002). Pre obe možno nájsť empirickú podporu.

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

## 5.2 Pohľad teórie komplexných systémov

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

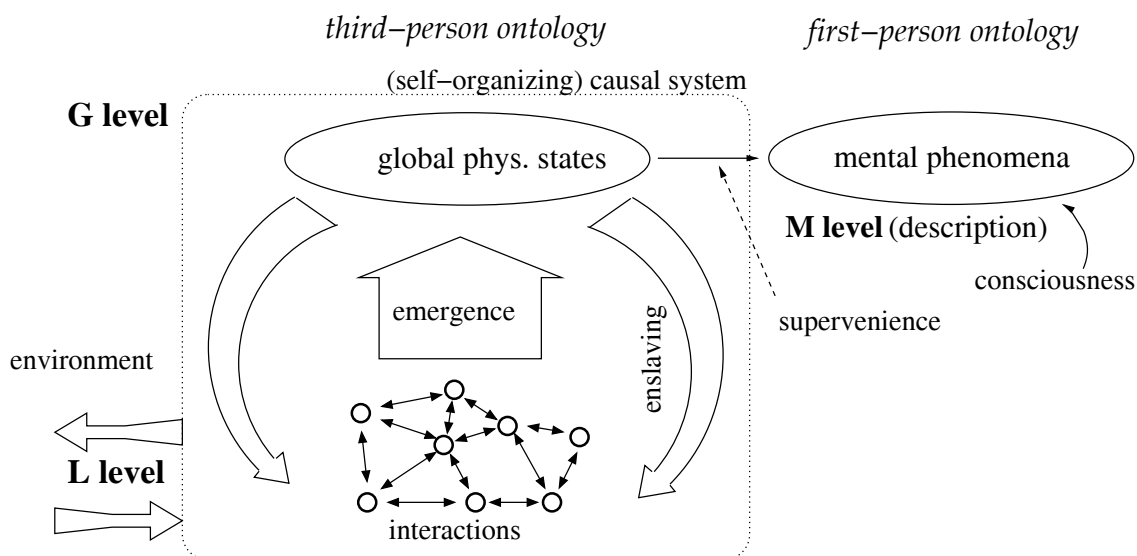
<sup>24</sup> Ako zlúčiť rôzne vlastnosti tej istej entity (napr. vnímaného objektu), aby vznikol koherentný vnem.

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

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

## 6 Úrovne opisu v modeli stelesnenej mysle

Vo Farkaš (2007) som načrtol konceptuálny rámec pre model stelesnenej mysle z pohľadu ponúknutého vysvetlenia recipročnej kauzality (podmienka E). Schéma najjednoduchšieho takéhoto modelu je na obr. 1. Model pozostáva z *troch* (namiesto filozoficky tradičných dvoch) úrovní opisu reality (dvoch fyzikálnych a jednej mentálnej), ktoré sú nevyhnutné na pochopenie podstaty recipročnej kauzality.<sup>26</sup> Kvôli zjednodušeniu považujeme za elementárnu fyzikálnu úroveň opisu biologické neuróny a za tú najvyššiu mozog ako celok. Medzi týmito okrajovými úrovňami existuje viacero fyzikálnych úrovní organizácie, ktoré sú relevantné z pohľadu neuroanatómie a neurofyziológie (neurálne mikroobvody, neurálne zhluky, či Brodmannove kortikálne arey), aj s ich vzájomnými interakciami (Churchland & Sejnowski, 1992), no v našom modeli budem od tohto faktu abstrahovať.



**Obrázok 1.** Náčrt formálneho modelu vzťahu myseľ-telo 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 1PP. Fyzikálny systém je otvorený, prijíma externé stimuly na úrovni L a spätne pôsobí na prostredie.

### 6.1 Fyzikálne úrovne opisu

Predpokladáme teda existenciu len dvoch interagujúcich „referenčných“ fyzikálnych úrovní, lokálnej L (časti mozgu) a globálnej G (mozog ako celok). Pritom treba zdôrazniť, že obe fyzikálne úrovne sú relatívne v tom zmysle, že fyzikálna mierka, ktorej sa budú týkať, závisí od kontextu. Napríklad, úroveň neurálnych zhlukov je globálna v kontexte úrovne neurónov, ale lokálna vo vzťahu ku kortikálnym areám.

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



Táto abstrakcia je užitočná pre pochopenie existujúcich empirických a teoretických argumentov v prospech recipročnej kauzality, ktoré sú relevantné pre rôzne priestorové mierky (Farkaš, 2007, časť 3).

Pozícia neredukcionistického fyzikalizmu by sa dala charakterizovať takto:

*Fyzikalizmus* – kauzálny systém je fyzikálny, a je kauzálne uzavretý.<sup>27</sup> Kauzálne vzťahy podliehajú fyzikálnym zákonitostiam na rôznych úrovniach (od molekulárnej po systémovú), pri zjednodušenom pohľade môžeme hovoriť o úrovni neurónov. Mentálne stavy sú závislé od fyzikálnej reality, t.j. ľudská myseľ nemôže existovať bez mozgu.

*Neredukcionizmus* – epistemologické poznanie mysle a jej úplné vysvetlenie nie je principiálne redukovateľné na opis jazykom fyzikálnej vedy (t.j. opis mozgu).<sup>28</sup>

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

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

Nevyhnutnou súčasťou modelu je participácia *vonkajšieho prostredia*. Mozog-myseľ si vyžaduje permanentnú interakciu s ním, a možno ešte dôležitejšie sú podnety, než samotné reakcie.<sup>29</sup> Vonkajšie prostredie je nevyhnutným zdrojom empirickej skúsenosti a z dynamického hľadiska je „partnerom“ v procese zjednávania. No kvôli fyzickej separácii od prostredia si myslím, že pri modelovaní ukotvenej mysle nemusíme zachádzať za hranice ľudského tela (ako navrhuje napr. Rockwell, 2007).

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

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

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

Mysleľ sa formuje v interakcii s prostredím, jedinečne u každého jedinca, no ukotvená ostáva v tele napriek tomu, že mentálne prejavy si vyžadujú interakciu s prostredím.

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

## 6.2 Mentálna úroveň opisu

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

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

(2) M je ako jediná prístupná bežnému jazyku a jediná umožňuje efektívnu verbálnu komunikáciu medzi subjektmi. Subjekt vlastne ani nemá na výber, pretože úroveň G (aktivita jeho vlastného tela-mozgu) mu je neprístupná. Súčasne tým umožňuje subjektu figurovať v kontexte populácie, ktorú môžeme chápať ako vyššiu úroveň organizácie s možnými komplexnými prejavmi kolektívneho správania. Aj keby sme v budúcnosti našli presné neurálne koreláty všetkých mentálnych stavov, ich opis by bol kvôli ich zložitosti prakticky nepoužiteľný (pozri aj súvis jazyka a vedomia v Beňušková, 2005 a v Dennett, 1996).

(3) Vedomé stavy súvisiace s niektorými stavmi v mozgu (časť stavov v M) sú možno emergentným fenoménom. Argument o zombiách (hypotetických tvoroch – ľudských kópiách s absenciou vedomia), podporujúci nevysvetliteľnosť „ťažkých“ problémov vedomia pomocou vedeckých metód je logicky možný, no možno nerealistický.<sup>31</sup> Sám Chalmers (1996) pripúšťa, že existencia zombií je málo pravdepodobná (s. 96). Z pohľadu nášho modelu vedomé stavy majú mentálnu kauzálnu silu prostredníctvom úrovne G, pričom si uvedomujeme príčinné vzťahy. Máme pocit, že *my* sa rozhodujeme, máme pocit slobodnej vôle. Súčasne však zrejme i nevedomé mentálne stavy majú kauzálnu silu („nadol“, resp. i v rámci kauzálnej domény), pričom sprievodné účinky sú mimo vedomia (človek si nie je vedomý príčin svojho rozhodovania). Absencia

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

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

participácie vedomia je typická pre rýchle konanie, napr. v športe (hokejista si až dodatočne uvedomí a precitne, aký dal krásny gól), ale možno aj pri pomalších procesoch. To preto, že sme pod vplyvom mnohých podnetov (mimo tela i v rámci neho), mozog všetky spracováva, no iba niektoré procesy sa dostanú na vedomú úroveň. (4) Vedomie (a teda špecifické stavy/procesy mozgu) sú možno žiaducim, evolučne vyvinutým fenoménom pre človeka ako sociálne a emocionálne založený živočíšny druh (Kováč, 2003). O emóciách som sa v príspevku vyjadroval len okrajovo, ale nepochybne limbický systém zohráva významnú úlohu v kognitívnych procesoch (LeDoux, 1996). Mohli sme ho chápať ako modulačný systém, ktorý na úrovni tela interaguje s neurálnym systémom.

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

## 7 Zhrnutie

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

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

## Literatúra

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

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

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

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

---

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

- Barsalou L.: Perceptual symbol systems. *Behavioral and Brain Sciences*, 22 (1999) 577-609.
- Bell A.: Levels and loops: The future of artificial intelligence and neuroscience. *Philosophical Transactions of the Royal Society London B*, 354, 2013-2020, 1999.
- Beňušková L.: Kde sa jazyk stretáva s vedomím. In: *Jazyk a kognícia*, Kalligram, Bratislava, 235-261, 2005.
- Berger H.: Über das Elektroenkephalogramm des Menschen. *Archiv für Psychiatrie und Nervenkrankheiten*, Berlin, 87 (1929) 527-570.
- Block N.: On a confusion about a function of consciousness. *Behavioral and Brain Sciences*, 18 (1995) 227-287.
- Burian J.: Samoorganizace a kognice. In: *Mysel', inteligencia a život*. Vydavateľstvo STU v Bratislave, 2007, 173-190.
- Calvin W.H., Ojemann G.A.: *Conversations with Neil's Brain. The Neural Nature of Thought and Language*, Addison-Wesley 1994.
- Clark A.: *Mindware: An Introduction to the Philosophy of Cognitive Science*. Oxford University Press 2001.
- Chalmers D.: *The Conscious Mind*. NY: Oxford University Press 1996.
- Chalmers D.: First-person methods in the science of consciousness. *Arizona Consciousness Bulletin*, 1999, <http://consc.net/papers/firstperson.html>.
- Churchland P.: *The Neurocomputational Perspective*. Cambridge, MA: Bradford Books 1989.
- Cleeremans A.: Principles of implicit learning. In D. Berry (Ed.) *How implicit is implicit learning?* Oxford: Oxford University Press (1997) 196-234.
- Cleeremans A.: Computational correlates of consciousness. *Progress in Brain Research* 150 (2005) 81-98.
- Crick F., Koch C.: Consciousness and neuroscience. *Cerebral Cortex*, 8 (1998) 97-107.
- Davidson D.: *Mental Events*. Reprinted in *Essays on Actions and Events*. Oxford: Clarendon Press 1970.
- Dehaene S., Naccache L., Cohen L., Bihan D., Mangin J., Poline J., Riviere D.: Cerebral mechanisms of word masking and unconscious repetition priming. *Nature Neuroscience*, 4 (2001) 752-758.
- Dehaene S., Naccache L.: Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. *Cognition* 79, (2001) 1-37.
- Dennett D.: *The Intentional Stance*. Cambridge, MA: MIT Press 1987.
- Dennett D.: *Kinds of Minds: Towards an Understanding of Consciousness* 1996. Preklad: Peregrin J.: Academia Praha 2004.
- Doya K.: What are the computations of the cerebellum, the basal ganglia, and the cerebral cortex? *Neural Networks*, 12 (1999) 961-974.
- Eguíluz V. a spol.: Scale-free brain functional networks. *Physical Review Letters* 94 (2005) 018102.
- Farkaš I.: Konekcionistické modelovanie jazyka. In: *Jazyk a kognícia*, Kalligram, Bratislava, 262-305, 2005.
- Farkaš I.: Samoorganizácia ako hybná sila dynamických vzorcov aktivít v mozgu a mysli. In: *Kognice a umělý život VI*, Opava: Slezská univerzita, 143-148, 2006.

Farkaš I.: Hľadanie kauzálnych vzťahov v probléme mysle a tela z pohľadu neredukcionistického fyzikalizmu. In: *Mysel', inteligencia a život*. Vydavateľstvo STU v Bratislave, 2007, 3-16.

Fodor J.: *The Modularity of Mind*. Cambridge, MA: MIT Press 1983.

Fodor J.: *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. Cambridge, MA: MIT Press 1987.

Fodor J., Pylyshyn Z.: Connectionism and cognitive architecture: A critical analysis. *Cognition*, 28 (1988) 3-71.

Friedenburg J., Silvermann G.: *Cognitive Science: An Introduction to the Study of Mind*. London, UK: Sage Publications 2006.

Gärdenfors P. Cued and detached representations in animal cognition. *Behavioural Processes*, 36 (1996) 263-273.

Griffin D., Speck G.: New evidence of animal consciousness. *Animal and Cognition*, 7 (2004) 5-18.

Havel I.: Přirozené a umělé myšlení jako filozofický problém. V knihe Mařík V. a kol.: *Umělá inteligence (3. díl)*. Academia Praha, 2001, 17-75.

Havel I.: Prožívání epizodických situací. In: Kvasnička V. a spol. (ed.): *Mysel', inteligencia a život*. Vydavateľstvo STU v Bratislave, 2007, 27-70.

Harnad S.: The symbol grounding problem. *Physica D*, 42 (1990) 335-346.

Hornik K., Stinchcombe M., White H.: Multilayer feedforward networks are universal approximations. *Neural Networks*, 2 (1989) 359-366.

Hurley S., Chater N. (Eds.): *Perspectives on Imitation: From Neuroscience to Social Science (Vol. 1, 2)*, Cambridge, MA: MIT Press, 2005.

Izhikevich E.: Which model to use for cortical spiking neurons? *IEEE Transactions on Neural Networks*, 15 (2004) 1063-1070.

Jedlička P., Beňušková L., Macáková J., Ostatníková D.: Molekulové mechanizmy učenia a pamäti. In: Hulín I. (ed.) *Patofyziológia* (6. vyd.) Slovak Academic Press, Bratislava, (2002) 1183-1199.

Johnson-Laird P.: *Mental Models*. Cambridge University Press 1983.

Hutchins E.: *Cognition in the Wild*. Cambridge, MA: MIT Press 1995.

Kaczmarek L., Levitan I. (eds): *Neuromodulation: The Biochemical Control of Neuronal Excitability*. NY: Oxford University Press 1987.

Kelso, J.A.S.: *Dynamic Patterns*. Cambridge, MA: MIT Press 1995.

Kim J.: *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*. Cambridge MA: MIT Press 1998.

Kosslyn S.: *Image and Brain: The Resolution of the Imagery Debate*. Cambridge, MA: MIT Press 1994.

Kováč, L.: Ľudské vedomie je produktom evolučnej eskalácie emocionálneho výberu. In: *Kognícia a umelý život III*. Elfa: Košice (2003) 21-39.

Kráľ A., Hulín I.: Neurofyziologické predpoklady vedomia. *Psychiatria*, 8 (2001) 86-98.

Kvasnička V.: Kognitívna veda, umelá inteligencia a teória ohraničenej racionality. In: *Kognice a umelý život VI*, Opava: Slezská univerzita, 233-252, 2006.

- LeDoux K.: *The Emotional Brain*. NY: Touchstone 1996.
- Marr D.: *Vision*. San Francisco, CA: W.H. Freeman 1982.
- Mayberry M.: Incremental nonmonotonic parsing through semantic self-organization. *Doctoral thesis*. Department of Computer Sciences, University of Texas at Austin 2003.
- Metzinger T. (ed.): *Neural Correlates of Consciousness: Empirical and Conceptual Questions*. Cambridge, MA: The MIT Press 2000.
- Newell A., Simon H.A.: Computer science as empirical enquiry. *Communications of the ACM*, 19 (1976) 113-126. In: *Mind Design II*, J. Haugeland (ed.), 81-110. Cambridge, MA: MIT Press 1997.
- O'Reilly R., Munakata Y.: *Computational Explorations in Cognitive Neuroscience: Understanding the Mind by Simulating the Brain*. MIT Press 2000.
- Pert C., Schmitt F.: Neuropeptides and their receptors: A psychosomatic network. *Journal of Immunology*, 135 (1985) 820-826.
- Port R., van Gelder T. (Eds.): *Mind as Motion: Explorations in the Dynamics of Cognition*. Cambridge, MA: MIT Press 1995.
- Pstružina K.: Abstrakce, komparace a generalizace jako základní operace lidského myšlení. In: *Mysel, inteligencia a život*. Vydavateľstvo STU v Bratislave, 2007, 111-117.
- Pylyshyn Z.: *Computation and Cognition*. MIT Press 1984.
- Ramsey W., Stich S., Garon, J.: Connectionism, eliminativism and the future of folk psychology. In: *Philosophy and Connectionist Theory*. Hillsdale, NJ: Lawrence Erlbaum (1991) 199-228.
- Rizzolatti G. Fadiga L., Gallese V., and Fogassi L.: Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, 3(2), (1996) 131-141.
- Rockwell T.: *Neither Brain nor Ghost: A Nondualist Alternative to the Mind-Brain Identity Theory*. Cambridge, Mass.: MIT Press, 2007.
- Rohde, D. A connectionist model of sentence comprehension and production. *Doctoral thesis*, Carnegie Mellon University, Pittsburg, PA 2002.
- Rohde D., Plaut D.: Connectionist models of language processing. *Cognitive Studies*, 10 (2003) 10-28.
- Searle J.: Consciousness. *Annual Reviews of Neuroscience* 23 (1999) 557-578.
- Seth A., Baars B., Edelman D.: Criteria for consciousness in humans and other mammals. *Consciousness and Cognition*, 14 (2005) 119-139.
- Seth A., Izhikevich E., Reeke G., Edelman G.: Theories and measures of consciousness: An extended framework. *Proceedings of the National Academy of Sciences*, 103 (2006) 10799-10804.
- Shafritz K., Gore J., Marois R.: The role of the parietal cortex in feature binding. *Proceedings of the National Academy of Sciences*, 99 (2002) 10917-10922.
- Siegelmann H., Sontag E.: Turing computability with neural nets. *Applied Mathematics Letters*, 4 (1991) 77-80.
- Singer W.: Neuronal synchrony: A versatile code for the definition of relations? *Neuron*, 24, (1999) 49-65.
- Spitzer M.: *The Mind within the Net: Models of Learning, Thinking and Acting*. Cambridge, MA: MIT Press 1999.

- Sporns O., Zwi J.: The small world of the cerebral cortex. *Neuroinformatics*, 2 (2004) 145-162.
- Sternberg R.: *Cognitive Psychology* (3. vyd.). Thomson Wadsworth 2003.
- Šefránek J.: Kognitívna teória usudzovania. In: *Mysel', inteligencia a život*. Vydavateľstvo STU v Bratislave, 2007, 233-239.
- Thelen E., Smith L.: *A Dynamic Systems Approach to the Development of Cognition and Action*. Cambridge, MA: MIT Press 1994.
- Tononi G., Sporns O., Edelman G.: A measure for brain complexity: relating functional segregation and integration in the nervous system. *Proceedings of the National Academy of Science*, 91 (1994) 5033-5037.
- van Gelder T.: Compositionality: A connectionist variation on a classical theme. *Cognitive Science*, 14 (1990) 355-384.
- van Gelder T., Port R.: It's about time: An overview of dynamical approaches to cognition. In: *Mind as Motion: Explorations in the dynamics of cognition*. Cambridge, MA: MIT Press, 1995, 1-44.
- Varela F.J., Thompson E., Rosch, E. *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge, MA: MIT Press 1991.
- Vavřečka M.: Ukotvení symbolů v kontextu nulového sémantického závazku. In: *Kognícia a umelý život VI*, Opava: Slezská univerzita, 2006, 401-411.
- Wiedermann J.: Mirror neurons, embodied cognitive agents and imitation learning. *Computing and Informatics*, 22 (2003) 545-559.
- Ziemke T.: What's that thing called embodiment? *Proceedings of the 25th Annual Meeting of the Cognitive Science Society*. Lawrence Erlbaum 2003.
- Zwaan R., Radvansky A.: Situation models in language comprehension and memory. *Psychological Bulletin*, 123 (1998) 162-185.