Je reprezentačný pluralizmus v kognitívnej vede nevyhnutný?

Igor Farkaš

Centrum pre kognitívnu vedu, Fakulta matematiky, fyziky a informatiky, Univerzita Komenského
Mlynská dolina, 84248 Bratislava
farkas@fmph.uniba.sk

Abstrakt

V príspevku sa zamýšľame nad vysvetľovaním povahy kognitívnych procesov ako prejavov mysle. Uvádzame a porovnávame existujúce teórie a paradigmy v kognitívnej vede z reprezentačno-výpočtového pohľadu, pričom zameriavame pozornosť na diskrétne a spojité formy reprezentácie, ako aj spojitú a diskrétnu dynamiku. Argumentujeme, že konekcionizmus má najbližšie k vysvetleniu mysle a jej prejavov, avšak nie z dôvodu spojitosti reprezentácií, ale vďaka schopnosti učenia a generalizácie. Pojednávame aj o abstrakcii ako dôležitom aspekte ľudskej kognície, ktorý objasňujeme z pohľadu neurovedy. V závere konštatujeme, že reprezentačný pluralizmus v kognitívnej vede je z epistemologického hľadiska užitočný, najmä v kontexte budovania inteligentných robotických systémov, no možno v budúcnosti nie nevyhnutný, čo sa týka mechanistického pochopenia mysle.

1 Úvod

Dlhodobou ambíciou kognitívnej vedy je vysvetliť podstatu ľudskej mysle ako entity, v ktorej prebiehajú kognitívne procesy subjektu. Primárne máme na mysli biologické subjekty (človeka alebo zvieratá), no môžeme uvažovať aj o (umelých) kognitívnych procesoch v súvislosti s artefaktami, napr. robotmi, od ktorých očakávame nejaké prejavy „inteligentného" správania. Komplikáciou v kognitívnej vede je to, že pojem mysle, na rozdiel od mozgu ako predpokladaného neurálneho substrátu, je vo svojej podstate vágny, a preto dáva priestor pre rôzne pohľady, v rámci ktorých sa ponúkajú odpovede. Napredujúce empirické vedy (psychológia, kognitívna neuroveda a iné) ovplyvňujú tieto paradigmy, pričom sa vychádza z konsenzu, že empirické poznatky sú zdrojom informácie pre modifikáciu vedeckých teórií, aj keď to sťažujú prípady nejednoznačnej interpretácie výsledkov experimentu.

Napriek komplikáciám sa ako dominantný prístup k poznávaniu mysle vyprofiloval reprezentačno-výpočtový pohľad [37] založený na hypotéze, že v mysli existujú *mentálne reprezentácie* (analógie dátových štruktúr), nad ktorými sa realizujú nejaké operácie (výpočty). Týmto sa však len presunulo bremeno vysvetľovania, pretože ani pojmy reprezentácie a počítania nie sú jednoznačne interpretované. Do akej miery tieto reprezentácie a výpočty súvisia s mozgom, závisí od paradigmy, ktorú uznávame. V každom prípade však môžeme predpokladať to, že kognitívne procesy prebiehajú v čase a v nejakom priestore reprezentácií. Potom sa môžeme pýtať, či ten čas a priestor sú vo svojej podstate diskrétne alebo spojité veličiny. Inak povedané, či je myseľ diskrétna alebo spojitá, alebo aj-aj. Reprezentačno-výpočtový prístup je akceptovaný v rôznych výpočtových paradigmách kognitívnej vedy (spomínaných ďalej). Výpočtové modelovanie považujeme za nepostrádateľnú súčasť mechanistického poznávania mysle, pretože okrem snahy o vysvetlenie ľudskej kognície tieto paradigmy možno využiť aj pri konštruovaní inteligentných (robotických) systémov, čím sa rozširuje horizont kognitívnej vedy [7].

Koexistencia rôznych vysvetlení kľúčových pojmov (reprezentácia, počítanie) vedie k určitému pluralizmu. V tomto príspevku sa zamýšľame nad otázkou pluralizmu z pohľadu umelej inteligencie (UI), ale aj psychológie a neurovedy. Ďalej budeme argumentovať, že reprezentačný pluralizmus je síce užitočný (v UI), no z hľadiska ontologickej podstaty ľudskej mysle nie nevyhnutný. Zdôvodníme, v čom tkvie výhoda konekcionistickej paradigmy, ktorá má spomedzi existujúcich prístupov najlepšie predpoklady na najkompletnejší opis kognitívnych procesov.

V tomto príspevku najprv stručne sumarizujeme hlavné teórie (časť 2) a hlavné výpočtové paradigmy v kognitívnej vede (časť 3), a tie potom navzájom porovnáme (časť 4). Ďalej zameriame pozornosť na porovnanie dvoch dynamík formálnych systémov (časť 5). Spomenieme zopár argumentov v kontexte rozdielností medzi symbolizmom a konekcionizmom (časť 6) a napokon pojednávame o neurovednom pohľade na abstrakciu ako dôležitú vlastnosť kognície (časť 7). Záver je stručným zhrnutím našej argumentácie (časť 8).

2 Teoretické smery v kognitívnej vede

Prvý teoretický smer – *symbolizmus* (kognitivizmus) bol podnietený zrodom konceptu Turingovho stroja ako hypo-

tetického univerzálneho počítača, a vynájdením moderných digitálnych počítačov. Základným konceptom sú diskrétne výpočty so *symbolmi*. Pod pojmom symbol máme na mysli *amodálny* symbol – konštrukt zavedený v kognitívnej psychológii na označenie reprezentácie nezávislej od modalít (vstupných senzorických podsystémov). Symbolový pohľad na myseľ je dodnes považovaný za tú správnu úroveň opisu najmä v kontexte vysvetľovania procesov tzv. vyššej kognície, ako je usudzovanie, plánovanie, a tiež používanie prirodzeného jazyka (napr. [8, 25, 30], kde kognitívne procesy sú oddelené od perceptuálnych a motorických procesov. Mentálne operácie sú realizované pomocou vnútorného jazyka mysle, napr. mentálčiny [8]. Výhodou symbolizmu je, že poskytuje silné matematické a logické formalizmy, ktoré sú zväčša transparentné a preto človeku zrozumiteľné. Pre človeka je prirodzené uvažovať v termínoch diskrétnych podmienok, pravidiel, propozícií a logických inferencií.

V kognitívnej psychológii sa v 70. rokoch minulého storočia začali objavovať skeptické pohľady na centrálnosť (vnútorných) reprezentácií a symbolového spracovania informácie v ľudskej kognícii a začali vznikať alternatívne teórie. K takýmto smerom patrí ekologická psychológia [10], ktorá zdôrazňovala, že „nie je dôležité, čo je vo vnútri hlavy, ale vnútri čoho (akého prostredia) sa hlava nachádza." Tento teoretický smer plynulo prešiel od 80. rokov do teórie situovanej kognície, ktorá bola postavená na predpokladoch, že ľudská inteligencia je fundamentálne interaktívna a fundamentálne neoddeliteľná od kontextu [13].

Dôraz na interakciu s prostredím ostal v centre pozornosti aj naďalej, no v 90. rokoch sa začal klásť dôraz na reprezentácie, ktoré túto interakciu umožňovali, a to v rámci *stelesnenej kognície*. Existuje viacero pohľadov na stelesnenú kogníciu [40]. Barsalou navrhol zastrešujúci termín *ukotvená kognícia* [2], ktorý zjednocuje rôzne aspekty stelesnenia a situovania v prostredí, vrátane konceptu tzv. rozšírenej mysle (extended mind). Silnú podporu pre túto paradigmu v kognitívnej vede možno nájsť v narastajúcej empirickej evidencii (za ostatných 20 rokov), či už z behaviorálnych štúdií alebo zobrazovacích metód (pozri referencie v [2]). Táto evidencia napovedá, že všetky kognitívne funkcie (vrátane tých vyšších) sa do istej miery „opierajú" o (nízkoúrovňové) senzomotorické procesy. Inými slovami, neurálne reprezentácie vyšších a nižších kognitívnych procesov majú prienik. Tieto teórie tiež pracujú s pojmom symbol, avšak v zmysle modálnych symbolov. Napríklad, vplyvná teória tzv. perceptuálnych symbolov [1] postuluje existenciu symbolov, ktoré sú realizované pomocou multimodálnych reprezentácií, a ktoré v podstate zodpovedajú pojmom (mentálnym kategóriám).

3 Výpočtové paradigmy v kognitívnej vede

V priebehu zhruba 60 rokov existencie kognívnej vedy sa v nej vyprofilovali štyri výpočtové paradigmy: symbolová, konekcionistická, dynamická a pravdepodobnostná[1] (v takomto poradí vzniku). Tieto paradigmy sa vzájomne líšia predpokladanými reprezentáciami a spôsobmi výpočtu.

Pojem reprezentácie sa bežne používa v kognitívnej psychológii, lingvistike i UI. Jazykový pôvod tohto pojmu napovedá, že ide o re-prezentovanie niečoho vonkajšieho (vo svete) niekde inde, vo vnútri nejakého systému (živého alebo umelého) alebo i na papieri. V kognitívnej psychológii sa hovorí o mentálnych reprezentáciách, čo je v podstate prvý teoretický konštrukt v kognitívnej vede [26]. Korene pojmu mentálna reprezentácia však siahajú až do antických čias, keď širšie chápanie tohto pojmu medzi filozofmi nemalo výpočtový charakter. Tradičná polemika v kognitívnej vede sa týka povahy týchto reprezentácií, čo reflektujú aj jednotlivé paradigmy kognitívnej vedy.

Symbolizmus je konceptuálne úzko spätý s digitálnym počítačom, ktorý realizuje diskrétne výpočty so symbolmi. Výpočet v počítači prebieha za pomoci dvoch kľúčových komponentov: procesora a pamäte. Interakcia počítača s prostredím je na periférii záujmu a prebieha prostredníctvom vstupno-výstupných podsystémov. Procesor sériovým spôsobom spracováva symboly, uložené v pamäti a vykonáva pritom inštrukcie podľa programu uloženého v inej časti pamäte. Klasická paradigma konceptualizuje myseľ ako výpočtový stroj, oddeliteľný od prostredia, ktorý manipuluje s internými symbolmi, odvodenými (zvonku) pomocou transdukcie z prostredia, podľa logických pravidiel. Výpočtovú teóriu mysle vystihuje hypotéza o fyzikálnom symbolovom systéme, ktorý „disponuje nutnými a postačujúcimi prostriedkami na všeobecné inteligentné konanie" [19]. Koncept univerzálneho Turingovho stroja je síce dôležitý v kontexte abstraktného počítania, avšak nie je relevantný pre UI a kognitívnu vedu, pretože neprispel k riešeniu praktických úloh UI, a ani k tomu, ako pracuje mozog [32].

Konekcionizmus spochybňuje základný predpoklad tradičnej UI, že mentálne procesy možno najlepšie charakterizovať ako algoritmické manipulácie so symbolmi. Konekcionizmus však nie je homogénnou skupinou, ale spektrom metód, ktoré boli inšpirované architektúrou a fungovaním mozgu (inšpirácia „zdola"). Sila konekcionistického systému – *umelej neurónovej siete* – nie je v samotných neurónoch, ale v ich vzájomných (excitačných a inhibičných) prepojeniach a interakcii. Paralelné spracovanie a

[1] V literatúre sa pravdepodobnostný prístup (verím, že zatiaľ) neobjavuje ako samostatná paradigma. Domnievam sa však, že je dostatočne principiálny a odlišný od ostatných prístupov, preto si nálepku paradigma zaslúži.

distribuovanosť aktivity[2] predstavujú základný rozdiel v architektúre v porovnaní so symbolovým systémom, pretože každý neurón je súčasne procesorom aj pamäťou (aj keď elementárnou). Taktiež, povaha komunikácie medzi neurónmi má numerický a nie symbolový charakter, preto hovoríme v prípade sietí o *subsymbolových* reprezentáciách. Neurónová sieť v podstate realizuje nelineárne vektorové operácie v metrickom (euklidovskom) priestore. Je zrejmé, že takéto operácie a reprezentácie sú v porovnaní so symbolovými oveľa menej transparentné (sieť ako „čierna skrinka"). Našťastie existujú techniky zhlukovania a vizualizácie mnohorozmerných dát, vďaka ktorým môžeme zisťovať, čo sa v neurónovej sieti deje [20].

Dynamický prístup je charakteristický úzkym prepojením agenta na prostredie (na rozdiel od symbolizmu a konekcionizmu, pozri aj [6]), v procese permanentnej vzájomnej interakcie, ktorá prebieha v spojitom priestore a čase, a ktorá sa dá najlepšie opísať pomocou diferenciálnych rovníc [27]. Dôraz sa teda kladie na situovanosť a stelesnenosť ľudského správania [29], čo je intuitívne zrejmé hlavne pri vysvetľovaní bezprostrednej senzomotorickej interakcie agenta s okolím. Dynamický prístup má tiež svoje vnútorné členenie. Radikálna dynamická platforma popiera akékoľvek reprezentácie, zatiaľ čo mäkšia platforma ich nevylučuje, čiže je konzistentná s reprezentačno-výpočtovým pohľadom. Dynamický prístup ku spoznávaniu vonkajšieho sveta je konzistentný so zjednávacím (enactive) prístupom [35], kde sa dôraz z vnútorných reprezentácií (vopred daného) vonkajšieho sveta presúva na vnímanie a jednanie vo svete, ktorý sa takto spoluvytvára. Vytráca sa dichotomické delenie na subjekt a objekt poznania.

Pravdepodobnostný prístup sa stal populárny v kognitívnej vede najmä v ostatnej dekáde rokov [23]. Ide o teoreticky podložený (bayesovský) prístup, ktorý využíva široké spektrum reprezentácií (stromy, vektory, logické pravidlá atď.), kombinuje ich so štatistickým učením a inferenciami za prítomnosti neurčitosti. Ponúka vysvetlenie rôznych prejavov ľudského správania (pozri [11] a tamojšie referencie). Postupuje zhora, od funkcie ktorú chceme vysvetliť, pričom sa hľadajú optimálne reprezentácie dát. Taktiež zahŕňa aspekt vrodené–získané v podobe tzv. induktívnych predispozícií (inductive biases), čo sú v podstate apriórne distribúcie veličín na množine hypotéz (vrodené), ktoré vstupujú do výpočtu posteriórnych distribúcií (získané). Jedným z problémov pravdepodobnostného prístupu je však jeho výpočtová neúnosnosť v prípade zložitejších problémov [18].

[2] Aj keď treba upresniť, že tzv. lokalistické modely neurónových sietí (teda nie distribuované) pracujú s reprezentáciami, ktoré možno nazvať symbolovými.

Úrovne analýzy	Symb	Kon	Dyn	Pr
výpočtová	+	+	+	+
algoritmická	+	+	+	+
implementačná	−	+	−	−

Tabuľka 1: Relevantnosť úrovní analýzy v jednotlivých výpočtových paradigmách kognitívnej vedy (symbolovej, konekcionistickej, dynamickej, a pravdepodobnostnej).

4 Porovnanie paradigiem

V snahe porovnať jednotlivé paradigmy sa na ne môžeme pozrieť z pohľadu teórie troch nezávislých úrovní analýzy (vysvetlenia nejakého fenoménu), a to výpočtovej, algoritmickej a implementačnej, ktorú Marr [17] rozpracoval v kontexte modelovania vizuálneho spracovania informácie, a ktorá výrazne ovplyvnila kognitívnu vedu. *Výpočtová úroveň* definuje výpočty, ktoré treba vykonať, napríklad pomocou matematickej funkcie alebo špecifikácie úlohy. *Algoritmická úroveň* špecifikuje použité reprezentácie uchovávajúce informácie, a výpočty s nimi. *Implementačná úroveň* špecifikuje procesy spracovania informácie, ktoré sú viazané na konkrétny hardvér, ktorý je použitý na implementáciu. Medzi úrovňami platí vzťah tzv. viacnásobnej realizovateľnosti, čo znamená, že jeden opis na vyššej úrovni sa dá transformovať na viacero opisov na nižšej úrovni.[3]

Marrova teória bola silne ovplyvnená symbolizmom, ktorý implementačnú úroveň považuje za nepodstatnú. Toto je pochopiteľné v prípade počítača, ktorý je naozaj duálnou entitou s nezávislým hardvérom a softvérom, kde algoritmy sa menia na implementáciu úplne automatickým procesom kompilácie (t. j. prekladu do hardvérovo-závislého strojového kódu). Avšak, v prípade mozgu nemožno neurálnu implementáciu automaticky odvodiť z nejakého opisu na vyššej úrovni. Mozog nebol nikým skonštruovaný, ale sa evolvoval tak, aby umožnil organizmu efektívne konať v dynamickom prostredí [3]. Implementačná úroveň hrá dôležitú úlohu a mala by byť vhodne spojená s opisom na vyššej úrovni, aby nám to uľahčilo interpretáciu jej funkcie.

Jednotlivé paradigmy možno porovnať, čo sa týka relevantnosti jednotlivých úrovní analýzy (viac v [7]). Porovnanie znázorňuje tabuľka 1. Je vidieť, že iba konekcionistická paradigma zahŕňa všetky tri úrovne. Symbolizmus a pravdepodobnostný prístup zjednocujú dva aspekty. Po prvé, oba prístupy ponúkajú vysvetlenia na výpočto-

[3] Ako ilustračný príklad uvažujme násobenie dvoch viaciferných čísel. To je teda cieľ výpočtu, ktorý sa dá dosiahnuť rôznymi algoritmami, napríklad takým, ktorý človek bežne používa, a v ktorom sa medzivýsledky násobenia (jedného čísla číslicou druhého čísla) zapíšu pod seba a potom sa sčítajú. Napokon, implementačná úroveň už predstavuje konkrétnu realizáciu tohto algoritmu v nejakom fyzickom médiu (počítač, pero a papier a i.).

	Učenie je silnou stránkou	
Distribuovanosť	–	+
–	Symbolizmus	Pravdep. modely
+	Dynam. systémy	Konekcionizmus

Tabuľka 2: Porovnanie paradigiem z pohľadu vybraných charakteristík.

vej úrovni, ktorú nepovažujú len za abstrakciu od inherentných mechanizmov (pravdepodobne neurálnych), ale za nezávislú úroveň vysvetlenia nejakého fenoménu. Po druhé, oba prístupy sú symbolové, no líšia sa v tom, že pravdepobnostné modely používajú spojité reprezentácie (pravdepodobností). Konekcionizmus a dynamické systémy predstavujú subsymbolové prístupy, pričom je medzi nimi súvis (rekurentné neurónové siete sú príkladmi dynamických systémov) [6, 24]. Porovnanie paradigiem z pohľadu uvedených charakteristík ponúkame v tabuľke 2. Tu máme na mysli to (1) či učenie je silnou stránkou paradigmy, a (2) či paradigma využíva lokalistické alebo distribuované reprezentácie.[4] Z tabuľky je vidieť, že aj z pohľadu týchto dvoch charakteristík sú neurónové siete pritipólom symbolizmu. Za kľúčovú vlastnosť neurónových sietí považujeme (1) *schopnosť zovšeobecnenia* (vďaka distribuovaným reprezentáciám) a (2) *schopnosť učenia*, t. j. realizáciu elementárnych zmien v znalosti systému (dlhodobej pamäti). V komplexných úlohách nie je principiálne možné vložiť všetky znalosti do systému; ten sa ich však môže naučiť (čo vo všeobecnosti nie je vôbec ľahké).

Aj keď každá z paradigiem má svoje špecifiká, najsignifikantnejší rozdiel sa týka *priepasti medzi symbolovými výpočtovými prístupmi a konekcionizmom* (kde sú kľúčové rekurentné neurónové siete schopné pracovať sekvenčne). Pravdepodobnostný prístup, ako symbolová inferencia v spojitom priestore, stojí niekde medzi nimi. Dynamické prístupy majú blízko ku konekcionizmu.

Symbolizmus a konekcionizmus predstavujú kvalitatívne odlišné alternatívy, aj niektorí autori vidia oba prístupy ako zlúčiteľné, figurujúce na rôznych úrovniach abstrakcie. Táto polemika sa tiahne de facto od 80. rokov minulého storočia (napr. [8, 33]). Oba tábory hovoria o vzájomnej nekompatibilite oboch prístupov, no na základe odlišných argumentov. Symbolisti tvrdia, že buď (1) oba prístupy sú nekompatibilné, alebo (2) konekcionistické modely sú len *implementáciou symbolových modelov* (implementačný konekcionizmus). Na druhej strane, podľa konekcionistov sú oba prístupy nekompatibilné preto, lebo *symbolové systémy nedokážu implementovať neurónové siete* vo všetkých prípadoch. Tento konekcionistický argument sa týka dynamiky systému, ktorú si vysvetlíme.

[4] Pre vysvetlenie dodajme, že pravdepodobnostné modely síce pracujú aj s distribuovanými reprezentáciami, avšak aktualizácia pravdepodobností hypotéz sa deje na symbolovej úrovni.

5 Subsymbolová a symbolová dynamika

V oboch paradigmách sa na kognitívne procesy pozeráme ako na deterministické sekvenčné procesy prebiehajúce v čase.[5] Symbolový systém spracováva symboly v čase podľa svojho programu, podobne ako dynamický systém sa vyvíja v stavovom priestore, hnaný aktivačnou dynamikou. Rozdiely medzi oboma paradigmami spočívajú v štruktúre stavového priestoru. Z hľadiska argumentácie nám stačí zamerať pozornosť na (stavový) priestor, v ktorom existujú reprezentácie (predpokladáme diskrétny čas). Základná otázka znie: Sú oba typy dynamík ekvivalentné, t. j. môžeme jednoznačne transformovať jednu na druhú? Alebo je symbolová dynamika abstrakciou, ktorá nie vždy dokáže opísať zložitosť dynamických dejov?

Deterministický dynamický systém možno opísať časovo-závislým stavovým vektorom $\mathbf{x}(t)$, ktorý sa vyvíja podľa deterministickej diferenčnej rovnice a vytvára tak trajektóriu (presnejšie povedané, množinu bodov v prípade diskrétneho času) v stavovom priestore X, čiže množine všetkých možných stavov. Napríklad, neurónová sieť s n sigmoidálnymi neurónmi na rekurentnej vrstve, t. j. s aktivačnou funkciou $f(z) = 1/(1+\exp(-z))$, má stavový priestor $X = [0,1]^n \subset R^n$, t. j. v tvare n-rozmernej kocky. V každom čase aktivita siete je reprezentovaná bodom v tomto stavovom priestore.

Je známe, že aj vo veľmi jednoduchom nelineárnom dynamickom systéme, môže vzniknúť veľmi zložité správanie (deterministický chaos), ktoré sa môže výrazne (kvalitatívne) meniť aj pri minimálnych zmenách parametrov. Príkladom je logistická mapa daná rovnicou $x(t+1) = r \cdot x(t) \cdot (1 - x(t))$, kde $x(t) \in (0,1)$ a $r \in \mathcal{R}$ je parameter. Hodnoty $x(t)$ sa ustália na rôznych hodnotách (nazývaných atraktory) práve v závislosti od r. Tieto ustálené hodnoty sa menia od jednej, cez dve, štyri hodnoty, až po chaotický režim v podobe tzv. podivných atraktorov (pri hodnote $r = 3.5699456...$). Chaotické systémy sú v prírode veľmi rozšírené a majú tú podstatnú vlastnosť, že ich vývoj v čase je v princípe nepredikovateľný v dlhšom horizonte, a to kvôli extrémnej závislosti od počiatočných podmienok (v literatúre známy ako „motýlí efekt").

Symbolová dynamika vznikne zo subsymbolovej tak, že stavový priestor rozdelíme na konečný počet neprekrývajúcich sa oblastí, ktorých zjednotenie tvorí celý priestor X. Vývoj systému potom môžeme opísať pomocou (konečnej) sekvencie symbolov, ktorá opisuje prechody medzi jednotlivými oblasťami. Napríklad v prípade logistickej mapy by sme interval (0,1) mohli rozdeliť napoly

[5] Teraz neuvažujeme stochastické modely neurónových sietí.

a prechody medzi oblasťami (alebo v rámci nich) opísať rôznymi symbolmi.

5.1 Sú obe dynamiky ekvivalentné?

Vzniká tu otázka, či opis na úrovni symbolovej dynamiky je rovnako presný a úplný ako ten na úrovni aktivačnej dynamiky. Inými slovami, ak sa rozhodneme opísať dynamický systém pomocou symbolovej dynamiky (použitím vhodnej partície), či sa nejaká informácia o vývoji systému stratí alebo nie. Tu je dôležitý koncept *topologickej ekvivalencie* medzi oboma priestormi. Dva priestory sú topologicky ekvivalentné práve vtedy, keď jeden môžeme spojito transformovať na ten druhý, a naopak (v matematike sa tomu hovorí homeomorfizmus).

Podmienkou pre zabezpečenie topologickej ekvivalencie je vytvorenie tzv. *generujúcej partície* (generating partition), ktorá má kľúčovú vlastnosť, že existuje jednojednoznačný vzťah medzi trajektóriami v stavovom priestore a sekvenciami symbolov (matematické vyjadrenie možno nájsť napr. v [12]). Problémom je to, že skonštruovať generujúcu partíciu vieme len v jednoduchších prípadoch. Doteraz bolo navrhnutých niekoľko algoritmov ako z pozorovaných dát (nemusíme poznať dynamiku, t. j. model systému) odhadnúť generujúcu partíciu, a to v prípade o niečo zložitejších chaotických dynamických systémov v 2D (pozri napr. [12] a tamojšie referencie). V prípade vysokorozmerných systémov však riešenia neexistujú. Z uvedeného teda vyplýva, že dynamické systémy sú výpočtovo silnejším formalizmom než symbolová dynamika. Subsymbolová dynamika má super-Turingovu silu, vďaka ktorej dokáže generovať dynamické správania, ktoré nie sú dosiahnuteľné pomocou symbolovej dynamiky [31, 36].

Avšak, položme si však otázku: *Je tento záver podstatný pre pochopenie mysle?* Niektorí sa nazdávajú, že áno, kvôli nutnosti vedieť vysvetliť niektoré „jemnôstky" pri učení, či kognitívnom vývine, napríklad pri akvizícii gramatiky jazyka [36]. Áno, potrebujme vysvetliť jednotlivé štádia kognitívneho vývinu, ale na to netreba nekonečnú presnosť v reprezentáciach. Toto naše tvrdenie podporíme niekoľkými argumentami:

(1) Neuróny pracujú v šume a vykazujú spontánnu aktivitu pálenia. Aj keď existujú rôzne teórie neurálneho kódu, na každom neuróne pravdepodobne nezáleží [28]. Tento neurálny šum nevnímame, a preto naša myseľ nepotrebuje nekonečnú presnosť.

(2) Introspektívne si človek uvedomuje len svoje mentálne stavy, ktoré sa menia oveľa pomalšie (rádovo stovky milisekúnd) než neurálne stavy (milisekundy). Okrem toho, vychádzame z predpokladu viacnásobnej realizovateľnosti.[6]

[6]Rôzne neurálne stavy znamenajú ten istý mentálny stav.

(3) Mozog pracuje zväčša v neautonómnom režime (prijíma externé vstupy a vysiela výstupy do prostredia), zatiaľ čo závery matematických dynamických modelov sa týkajú autonómneho režimu. Analýza neautonómnych systémov je nesmierne obtiažna. Viacero prác z oblasti modelovania komplexných systémov prichádza s hypotézou, že činnosť mozgu sa pohybuje na hranici chaosu (on the edge of chaos), vďaka čomu si pravdepodobne zachováva obrovskú komplexnosť a variabilnosť, ktorú má ešte poznávajúci subjekt pod kontrolou (napr. [15]).

6 Ešte zopár argumentov

Napriek tomu, že pomocou symbolových výpočtov by sme vedeli opísať (z hľadiska požadovanej presnosti) všetky dynamické deje, je užitočnejšie uvažovať o spojitej mysli, pretože spojitosť možno z matematického hľadiska vnímať ako všeobecnejší prípad.[7] V prospech poznávania spojitej mysle (ako dynamického systému) hovoria aj empirické poznatky z kognitívnej psychológie [34].

Na oba prístupy sa možno pozrieť aj z perspektívy úrovne opisu. Skok od spojitého konekcionizmu k diskrétnemu symbolizmu však znamená zmenu kvality. V literatúre sa objavujú argumenty v prospech ekvivalencie medzi oboma úrovňami, preto k nim zaujmeme stanovisko.

Argument 1: Dynamické systémy možno simulovať na digitálnom počítači. Áno, všetko, čo sa dá formalizovať (algoritmizovať) vieme simulovať na digitálnom počítači s vysokou presnosťou (s výnimkou chaotickej dynamiky, kde by sme potrebovali nekonečnú presnosť počítača). Na počítači vieme simulovať diskrétne aj spojité procesy, to však nič nehovorí o ontologickej podstate týchto procesov. Počítač tu figuruje len ako simulačný prostriedok.

Argument 2: Dynamické systémy možno opísať pomocou výpočtov. K odpovedi na tento argument je nutné objasniť pojem počítania. Klasické počítanie sa týka diskrétnych systémov, neklasické počítanie zahŕňa aj systémy so spojitým priestorom stavov (napr. neurónové siete so sigmoidálnymi aktivačnými funkciami). Dynamické systémy sú teda príkladom neklasických výpočtov, ktoré vieme simulovať pomocou diskrétnych výpočtov. Viac o počítaní možno nájsť v prácach [24, 7].

Argument 3: Ľudská myseľ je tak komplexná, že je nutné oprieť sa o reprezentačný pluralizmus [5]. Nazdávam sa, že pri súčasnom stave poznania je pluralizmus najschodnejšou cestou (hybridné systémy sú v UI dominantné [38]). Z evolučného a neurovedeného hľadiska za pravdepodobnejšie považujeme jednotné vysvetlenie, napríklad v duchu tzv. kognitívnej inkrementálnosti [4].

[7]Na spracovanie diskrétnych dát im stačí v neurónovej sieti priradiť symbolové, lokalistické reprezentácie, kde v každej reprezentácii je len jeden neurón aktívny.

Argument 4: Aj v iných vedách, napríklad vo fyzike sa nakoniec dospelo k viacerým vysvetleniam (napr. Newtonova fyzika a kvantová fyzika). Prečo by to tak nemohlo byť aj v prípade kognitívnej vedy? Na rozdiel od fyziky, ktorá vysvetľuje podstatu reality týkajúcu sa rôznych priestorových škál, sa v prípade mozgu nezdá, že máme do činenia s tak rozdielnymi škálami. Áno, môžeme hovoriť o molekulárnej úrovni na jednej strane a systémovej úrovni na strane druhej. Nazdávam sa, že vysvetlenie kognitívnych procesov si vyžaduje systémovú úroveň (napr. predpokladané kvantové fenomény v mozgu považujeme za irelevantné pre vysvetlenie mysle; pozri aj [16]),

7 Abstrakcia z pohľadu neurovedy

Na spektrum ľudských schopností sa môžeme pozrieť aj z pohľadu abstrakcie. Vysokoúrovňové procesy ako napríklad rozmýšľanie o matematickej rovnici vnímame ako viac abstraktné, než uchopenie pohára do ruky. Abstraktné veci sa bežne spájajú s využitím symbolov, zatiaľ čo konkrétne veci už menej. Tradičný pohľad v kognitívnej vede je typicky dichotomický. Ako sa na abstrakciu pozerá neuroveda a konekcionizmus? Pojem abstrakcie má viacero príbuzných významov. V našom kontexte môžeme povedať, že abstrakcia je daná *invariantnosťou*. Zhruba povedané: čím je reprezentácia aktivovateľná väčším spektrom vstupných podnetov, tým je abstraktnejšia. Abstrakcia je kľúčová vlastnosť kognície. Umožňuje nám kategorizovať vstupy, abstrahujúc od detailov, všímaním si len podstatných charakteristík.

Symbolový a subsymbolový pohľad môžeme vnímať ako stojace na opačných koncoch pomyslenej dimenzie abstrakcie, ktorú považujeme za kontinuum. Subsymbolové reprezentácie sú distribuované, existujú v podobe numerických vektorov v euklidovskom priestore s definovanou metrikou. Sú aktivované vtedy, keď vstupný podnet je dostatočne podobný (v zmysle danej metriky). Symbolové reprezentácie stoja na vrchole abstrakcie, sú kategorické. Sú aktivované pri širokom spektre vstupných podnetov, čiže sú vysoko invariantné voči charakteristikám podnetu.

Ako tieto dva odlišné svety premostiť? Človek predsa dokáže pracovať s oboma typmi entít. Ako je to s pojmami? V ľudskej kognícii ich existuje celé spektrum, od veľmi konkrétnych (napr. pes) až po vysoko abstraktné (napr. demokracia), ktoré si človek dokáže osvojiť, aj keď nie naraz, ale v postupných štádiách vývinu. Je známe, že dieťa si najprv osvojuje konkrétne veci a až neskôr abstraktnejšie, čo je podmienené maturačnými procesmi prebiehajúcimi v mozgu. Abstrakcia sa týka nielen sveta objektov ale aj priestoru udalostí či akcií. Rozhodnutie čoho sa napiť je zrejme jednoduchšie (konkrétnejšie) než rozhodnutie, na akú vysokú školu sa prihlásiť.

Neuróny v mozgu figurujú na rôznych vrstvách hierarchie. Majú však spoločné to, že každý neurón spracováva nejaký mnohorozmerný vstup a v závislosti od svojich parametrov (synaptických váh) reaguje naň alebo nie. Narastajúca abstrakcia, čo sa týka aktivovania neurónu súvisí s jeho invariantnosťou voči niektorým charakteristikám vstupov (ktoré tiež môžu byť viac alebo menej abstraktné). Typickou vlastnosťou mozgu sú asociačné oblasti, kde dominujú *multimodálne neuróny*, ktoré reagujú na podnety z viac ako jednej modality. Abstrakcia vzniká pomocou kaskády nelineárnych operácií medzi vrstvami. Na rozdiel od amodálnych symbolov (v symbolových systémoch), tento princíp *nevyžaduje proces transdukcie a abstrakcia vzniká postupne v procese učenia*.

Existujú početné poznatky o organizácii mozgu, čo sa týka invariantnosti neurónov pri reakcii na rôzne podnety. Napríklad vizuálny systém cicavcov je organizovaný hierarchicky v zmysle extrakcie príznakov (od jednoduchších ku zložitejším), čo sa dosahuje zväčšovaním rádiusu receptívneho poľa neurónov (aferentných spojení) smerom k vyšším vrstvám. Na vrchole pyramídy stojí area IT (inferior temporal) ako oblasť invariantého rozpoznávania objektov (voči pozícii, rotácii, či škále, a v prípade biologických objektov i deformácie) [14]. Podobne area STS (Superior Temporal Sulcus) má svoje členenie v kontexte rozpoznávania biologických pohybov.

Analogicky je to v prípade exekutívnej funkcie mozgu – konania. Predný lalok je hierarchicky organizovaný (od primárnej motorickej kôry smerom k prefrontálnej kôre), pričom jednotlivé časti sa podieľajú na rozhodovacích procesoch s rôznym stupňom abstrakcie [21]. Najvyššie v pyramíde stojí predná časť prefrontálnej kôry. Pozoruhodné je aj to, že jednotlivé časti frontálnej kôry majú analogickú štruktúru (projenie s talamom a niektorým z jadier v bazálnych gangliách, ktoré sa podieľajú na výbere akcie [22]. Ukazuje sa teda, že podobné (alebo rovnaké) neurálne mechanizmy operujú na rôznych úrovniach hierarchie (abstrakcie).

Aj organizácia dlhodobej pamäti v mozgovej kôre má hierarchickú organizáciu [9], v rámci ktorej existujú vzájomné prepojenia v prednej a zadnej časti kôry umožňujúce vybrať multimodálny obsah a aktivovať ho v podobe pracovnej pamäti.

Ako si teda ľudský mozog poradí s rôznymi entitami, čo sa týka miery abstrakcie? Je známe, že kôra má pomerne homogénnu štruktúru (kortikálne stĺpce) naprieč lalokmi, aj keď pravdepodobne existujú rozdiely v distribuovanosti reprezentácií v jednotlivých lalokoch mozgu. Ako argumentujú O'Reilly a spol. [22], v posteriórnej kôre sú neurálne reprezentácie viac distribuované, čo umožňuje vznik multimodálnych asociácií a dobrej generalizácie, zatiaľ čo vo frontálnej kôre sú reprezentácie riedke (sparse), aby sa zabránilo nežiaducim interferenciám (napr. pri aktivivaní viacerých alternatív pri rozhodovaní). Riedke reprezentá-

cie existujú aj v hipokampe, ako centre podieľajúcom sa na tvorbe epizodickej pamäti.

Neurovedné poznatky sú využiteľné pri modelovaní abstrakcie pomocou umelých neurónových sietí. Tento opis bude kvalitatívne rovnaký či už ide o psa alebo demokraciu. Všetky koncepty sú v mozgu reprezentované podobne ako vzorce aktivít, s rôznou mierou distribuovanosti, a v rôznych častiach mozgu [39]. Preto aj odpovedajúce konekcionistické reprezentácie budú mať podobné vlastnosti, v rámci jednotného prístupu. O niečo zložitejšie budú výpočtové operácie nad týmito reprezentáciami, v snahe modelovať fenomény ľudskej kognície ako napríklad kompozicionalita, systematickosť a problém viazania (pozri [7]). Tieto smery výskumu ostávajú predmetom záujmu.

8 Záver

Na základe použitej argumentácie dospievame k týmto záverom: Najsľubnejším prístupom k vysvetleniu mechanizmov mysle je konekcionistický prístup. Nie preto, že umožňuje realizovať neklasické výpočty (nekonečná presnosť nie je nevyhnutná), ale pretože umožňuje modelovať elementárne mechanizmy učenia, ktoré sú v kognícii kľúčové, pričom tieto mechanizmy nie sú nezávislé od implementančnej úrovne. Okrem toho, konekcionizmus ponúka možnosti modelovania kognitívnych procesov na rôznych úrovniach abstrakcie jednotným spôsobom (pomocou rovnakých princípov). Ako odpoveď na otázku v článku tvrdíme, že existujúci reprezentačný pluralizmus je síce užitočný (najmä v umelej inteligencii, kde sa nemusíme opierať o biologicky inšpirované riešenia), no možno sa ukáže v budúcnosti ako nie nevyhnutný pre výpočtovú kognitívnu vedu.

Poďakovanie

Tento príspevok bol podporený grantami VEGA 1/0439/11 a 1/0503/13, a KEGA 076UK-4/2013. Ďakujem svojim kolegom Jánovi Rybárovi a Martinovi Takáčovi za užitočné pripomienky.

Literatúra

[1] Barsalou, L.: Perceptual symbol systems. Behavioral and brain sciences 22(04), 577–660 (1999)

[2] Barsalou, L.: Grounded cognition. Annual Reviews of Psychology 59, 617–645 (2008)

[3] Churchland, P.S., Sejnowski, T.J.: The Computational Brain. The MIT Press (1992)

[4] Clark, A.: Mindware: An Introduction to the Philosophy of Cognitive Science. Oxford University Press (2001)

[5] Dove, G.: Beyond perceptual symbols: A call for representational pluralism. Cognition 110, 412–431 (2009)

[6] Farkaš, I.: Konekcionizmus v náručí výpočtovej kognitívnej vedy, pp. 19–62 (2011)

[7] Farkaš, I.: Indispensability of computational modeling in cognitive science. Journal of Cognitive Science 13(12), 401–435 (2012)

[8] Fodor, J.: The Mind Doesn't Work That Way. MIT Press (2000)

[9] Fuster, J.: Cortex and memory: Emergence of a new paradigm. Journal of Cognitive Neuroscience 21(11), 2047–2072 (2009)

[10] Gibson, J.: Ecological Approach to Visual Perception. Houghton-Mifflin, Boston (1970)

[11] Griffiths, T., Chater, N., Kemp, C., Perfors, A., Tenenbaum, J.: Probabilistic models of cognition: Exploring representations and inductive biases. Topics in Cognitive Science 14, 357–364 (2010)

[12] Hirata, Y., Judd, K., Kilminster, D.: Estimating a generating partition from observed time series: Symbolic shadowing. Physics Review E 70, 016215 (2004)

[13] Hutchins, E.: Cognition in the Wild. MIT Press (1996)

[14] Jellema, T., Perrett, D.: Neural representations of perceived bodily actions using a categorical frame of reference. Neuropsychologia 44, 1535–1546 (2006)

[15] Kitzbichler, M., Smith, M., Christensen, S., Bullmore, E.: Broadband criticality of human brain network synchronization. PLoS Computational Biology 5(3) (2009)

[16] Litt, A., Eliasmith, C., Kroon, F., Weinstein, S., Thagard, P.: Is the brain a quantum computer? Cognitive Science 30, 593–603 (2006)

[17] Marr, D.: Vision. W.H. Freeman, San Francisco, CA (1982)

[18] McClelland, J.: The place of modeling in cognitive science. Topics in Cognitive Science 1(1), 11–38 (2009)

[19] Newell, A., Simon, H.: Computer science as empirical enquiry. Communications of the ACM 19, 113–126 (1976)

[20] O'Brien, G., Opie, J.: How do connectionist networks compute? 7(1), 30–41 (2006)

[21] O'Reilly, R.: The what and how of prefrontal cortical organization. Trends in Neurosciences 33(4), 355–361 (2010)

[22] O'Reilly, R., Munakata, Y., Frank, M., Hazy, T., Contributors: Computational Cognitive Neuroscience. Wiki Book, second edn. (2012)

[23] Perfors, A., Tenenbaum, J., Griffiths, T., Xu, F.: A tutorial introduction to Bayesian models of cognitive development. Cognition 120, 302–321 (2011)

[24] Piccinini, G.: Some neural networks compute, others don't. Neural Networks 21(23), 311–321 (2008)

[25] Pinker, S.: How the Mind Works. W. W. Norton and co. (2009)

[26] Pitt, D.: Mental Representation (2012), http://plato.stanford.edu/archives/win2012/entries/-mental-representation

[27] Port, R., van T. van Gelder (eds.): Mind as Motion: Explorations in the Dynamics of Cognition. MIT Press, Cambridge, MA (1995)

[28] Rieke, F., Warland, D., de Ruyter van Steveninck, R., Bialek, W.: Spikes: Exploring the Neural Code. MIT Press (1999)

[29] Schoener, G.: Dynamical systems approaches to cognition. In: Sun, R. (ed.) Cambridge Handbook of Computational Psychology, pp. 101–126. Cambridge University Press, New York (2008)

[30] Šefránek, J.: Inteligencia ako výpočet. IRIS Bratislava (2002)

[31] Siegelmann, H.: Neural and super-turing computing. Minds and Machines 13(1), 103–114 (2003)

[32] Sloman, A.: The irrelevance of Turing machines to artificial intelligence. In: Scheutz, M. (ed.) Computationalism: New Directions, pp. 87–128. MIT Press, Cambridge (2002)

[33] Smolensky, P.: On the proper treatment of connectionism. Behavioral and Brain Sciences 11, 1–23 (1988)

[34] Spivey, M.: The Continuity of Mind. Oxford University Press, Oxford (2007)

[35] Stewart, J., Gapenne, O., Di Paolo, E. (eds.): Enaction: Toward a New Paradigm for Cognitive Science. MIT Press (2010)

[36] Tabor, W.: A dynamical systems perspective on the relationship between symbolic and nonsymbolic computation. Cognitive Neurodynamics 3(4), 415–427 (2009)

[37] Thagard, P.: Mind: An Introduction to Cognitive Science. MIT Press, second edn. (2005)

[38] Vernon, D., Metta, G., Sandini, G.: A survey of artificial cognitive systems: Implications for the autonomous development of mental capabilities in computational agents. IEEE Transactions on Evolutionary Computation 11(2), 151–180 (2007)

[39] Wang, J., Conder, J., Blitzer, D., Shinkareva, S.: Neural representation of abstract and concrete concepts: A meta-analysis of neuroimaging studies. Human Brain Mapping 31(10), 1459–1468 (2010)

[40] Wilson, M.: Six views of embodied cognition. Psychonomic Bulletin and Review 9(4), 625–636 (2002)